

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Jan 11 15:44:58 2000; MasPar time 5.30 Seconds
205.769 Million cell updates/sec

Search output not generated.

Title: >US-09-290-049-2
Description: (1-20) From US09290049.pep
Perfect Score: 135
Sequence: 1 VPSYSFIRTAHSEVDLIA 20

Scoring table: PAM 150
Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:

sptrembl9
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mmc 8:sp-organelle
9:sp-phase 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 28.249; Variance 35.475; scale 0.796

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	111	82.2	1390	2	069385	GLUCOSYLTRANSFERASE-SI	2.83e-12
2	111	82.2	1455	2	069388	GLUCOSYLTRANSFERASE-SI	2.83e-12
3	111	82.2	1455	2	069382	GLUCOSYLTRANSFERASE-SI	2.83e-12
4	111	82.2	1455	2	069397	GLUCOSYLTRANSFERASE-SI	2.83e-12
5	111	82.2	1455	2	069391	GLUCOSYLTRANSFERASE-SI	2.83e-12
6	111	82.2	1476	2	069390	GLUCOSYLTRANSFERASE-SI	2.83e-12
7	111	82.2	1476	2	069387	GLUCOSYLTRANSFERASE-SI	2.83e-12
8	111	82.2	1476	2	069384	GLUCOSYLTRANSFERASE-SI	2.83e-12
9	111	82.2	1476	2	069381	GLUCOSYLTRANSFERASE-SI	2.83e-12
10	111	82.2	1476	2	069396	GLUCOSYLTRANSFERASE-SI	2.83e-12
11	111	82.2	1476	2	069393	GLUCOSYLTRANSFERASE-SI	2.83e-12
12	111	82.2	1476	2	069390	GLUCOSYLTRANSFERASE-SI	2.83e-12
13	111	82.2	1476	2	069387	GLUCOSYLTRANSFERASE-SI	2.83e-12
14	111	82.2	1476	2	069384	GLUCOSYLTRANSFERASE-SI	2.83e-12
15	111	82.2	1476	2	069381	GLUCOSYLTRANSFERASE-SI	2.83e-12
16	111	82.2	1476	2	069396	GLUCOSYLTRANSFERASE-SI	2.83e-12
17	111	82.2	1476	2	069393	GLUCOSYLTRANSFERASE-SI	2.83e-12
18	111	82.2	1476	2	069390	GLUCOSYLTRANSFERASE-SI	2.83e-12
19	111	82.2	1476	2	069387	GLUCOSYLTRANSFERASE-SI	2.83e-12
20	111	82.2	1476	2	069384	GLUCOSYLTRANSFERASE-SI	2.83e-12

21	79	58.5	1462	2	069389	GLUCOSYLTRANSFERASE-S	2.52e-04
22	78	57.8	575	5	P90900	CYTOLASTIC INTERMEDI	4.28e-04
23	77	57.0	1518	2	000600	GLUCOSYLTRANSFERASE I	7.24e-04
24	75	55.6	1577	2	054178	GLUCOSYLTRANSFERASE	2.05e-03
25	74	54.8	1599	2	000599	GLUCOSYLTRANSFERASE	3.42e-03
26	72	53.3	101	3	006139	CHROMOSOME XII COSMID	9.49e-03
27	71	52.6	810	6	077788	NEUROFILAMENT-M SUBUN	1.57e-02
28	71	52.6	845	11	063370	NEUROFILAMENT PROTEIN	1.57e-02
29	71	52.6	1577	2	055265	GLUCOSYLTRANSFERASE PR	1.57e-02
30	70	51.9	494	4	016352	NEUROFILAMENT-66.	2.59e-02
31	70	51.9	501	11	061958	(CLONE PAMF-66).	2.59e-02
32	69	51.1	411	3	060057	HYPOTHETICAL 46.2 KD P	4.25e-02
33	68	50.4	490	13	P79933	HEPITIN.	6.94e-02
34	67	49.6	798	13	090307	HYPOTHETICAL 90.0 KD P	1.13e-01
35	66	48.9	913	13	013099	MIDDLE MOLECULAR WEIGH	1.83e-01
36	65	48.1	472	13	007962	GELFILTEN.	2.95e-01
37	65	48.1	897	13	013098	MIDDLE MOLECULAR WEIGH	2.95e-01
38	65	48.1	1110	13	091255	NE-180.	2.95e-01
39	64	47.4	342	5	025031	CYSTINE PROTEINASE.	4.73e-01
40	63	46.7	471	13	P87360	HYPOTHETICAL 86.2 KD P	7.56e-01
41	63	46.7	771	2	P71602	CYTOSOLIC P450 (P450)	1.20e+00
42	62	45.9	132	5	025509	T-CLUSTER BINDING PROT	1.20e+00
43	62	45.9	265	4	015187	HYPOTHETICAL 33.6 KD P	1.20e+00
44	62	45.9	287	2	051618		
45	61	45.2	290	2	083954		

ALIGNMENTS

RESULT	ID	069385	PRELIMINARY:	PRT:	1390 AA.
AC	069385	1			
DT	01-AUG-1998	(TREMBLREL. 07, CREATED)			
DT	01-AUG-1998	(TREMBLREL. 07, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	GLUCOSYLTRANSFERASE-SI.				
GN	GTFC.				
OS	STREPTOCOCCUS MUTANS.				
OC	BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;				
OC	STREPTOCOCCUS.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-MT4245;				
RX	MEDLINE: 98231643.				
RA	FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,				
RA	KIMURA S., HAMADA S.,				
RT	"Molecular analyses of glucosyltransferase genes among strains of				
RT	Streptococcus mutans."				
RT	FEMS MICROBIOL. LETT. 161:331-336(1998).				
DR	EMBL: D88655; D1027046; -				
DR	TRANSFERASE.				
SO	SEQUENCE 1390 AA; 155375 MW; C211E7B8 CMC32;				

Query Match
Best Local Similarity 94.7%; Pred. No. 2.83e-12;
Matches 18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Db	578 VPSYSFIRTAHSEVDLI 595
Qy	1 VPSYSFIRTAHSEVDLI 19
RESULT	2
ID	069388
AC	069388
DT	01-AUG-1998 (TREMBLREL. 07, CREATED)
DT	01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE	GLUCOSYLTRANSFERASE-SI.
GN	GTFC.
OS	STREPTOCOCCUS MUTANS.
OC	BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC	STREPTOCOCCUS.

RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN-MT4239;
RX MEDLINE: 98231643.
RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
RA KIMURA S., HAMADA S.,
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS MICROBIOL. LETT. 161:331-336(1998).
DR EMBL: D88658; D1027050; -.
KM TRANSFERASE.
SQ SEQUENCE 1455 AA; 163046 MW; 7659CPRC CRC32;

Query Match 82.2%; Score 111; DB 2; Length 1455;
Best Local Similarity 94.7%; Pred. No. 2.83e-12;
Matches 18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

DB 578 VPSTSFIR-AHDSEVODLI 595
OY 1 VPSTSFIRTAHDSEVODLI 19
|||||

RESULT 3 PRELIMINARY; PRT; 1455 AA.
ID 069382;
AC 069382;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GLUCOSYLTRANSFERASE-SI.
GN GTF.
OS STREPTOCOCCUS MUTANS.
OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MT8148;
RX MEDLINE: 98231643.
RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
RA KIMURA S., HAMADA S.,
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS MICROBIOL. LETT. 161:331-336(1998).
DR EMBL: D88652; D1027042; -.
KM TRANSFERASE.
SQ SEQUENCE 1455 AA; 162970 MW; 40B022BD CRC32;

Query Match 82.2%; Score 111; DB 2; Length 1455;
Best Local Similarity 94.7%; Pred. No. 2.83e-12;
Matches 18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

DB 578 VPSTSFIR-AHDSEVODLI 595
OY 1 VPSTSFIRTAHDSEVODLI 19
|||||

RESULT 4 PRELIMINARY; PRT; 1455 AA.
ID 069397;
AC 069397;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GLUCOSYLTRANSFERASE-SI.
GN GTF.
OS STREPTOCOCCUS MUTANS.
OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MT4467;
RX MEDLINE: 98231643.
RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
RA KIMURA S., HAMADA S.,
RT "Molecular analyses of glucosyltransferase genes among strains of

RT Streptococcus mutans.";
RL FEMS MICROBIOL. LETT. 161:331-336(1998).
DR EMBL: D89978; D1027060; -.
KM TRANSFERASE.
SQ SEQUENCE 1455 AA; 162914 MW; 7D729DA0 CRC32;

Query Match 82.2%; Score 111; DB 2; Length 1455;
Best Local Similarity 94.7%; Pred. No. 2.83e-12;
Matches 18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

DB 578 VPSTSFIR-AHDSEVODLI 595
OY 1 VPSTSFIRTAHDSEVODLI 19
|||||

RESULT 5 PRELIMINARY; PRT; 1455 AA.
ID 069391;
AC 069391;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GLUCOSYLTRANSFERASE-SI.
GN GTF.
OS STREPTOCOCCUS MUTANS.
OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MT4251;
RX MEDLINE: 98231643.
RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
RA KIMURA S., HAMADA S.,
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS MICROBIOL. LETT. 161:331-336(1998).
DR EMBL: D88661; D1027054; -.
KM TRANSFERASE.
SQ SEQUENCE 1455 AA; 162804 MW; DB730514 CRC32;

Query Match 82.2%; Score 111; DB 2; Length 1455;
Best Local Similarity 94.7%; Pred. No. 2.83e-12;
Matches 18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

DB 578 VPSTSFIR-AHDSEVODLI 595
OY 1 VPSTSFIRTAHDSEVODLI 19
|||||

RESULT 6 PRELIMINARY; PRT; 1476 AA.
ID 069390;
AC 069390;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GLUCOSYLTRANSFERASE-SI.
GN GTF.
OS STREPTOCOCCUS MUTANS.
OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MT4251;
RX MEDLINE: 98231643.
RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
RA KIMURA S., HAMADA S.,
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS MICROBIOL. LETT. 161:331-336(1998).
DR EMBL: D88660; D1027053; -.
KM TRANSFERASE.
SQ SEQUENCE 1476 AA; 165833 MW; 9119C72B CRC32;

Query Match 82.2%; Score 111; DB 2; Length 1476;

Best Local Similarity 94.7%; Pred. No. 2.83e-12;
Matches 18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Db 552 VPYSFIR-AHSEVODLI 569
OY 1 VPYSFIRTAHSEVODLI 19

RESULT 7
ID 069387 PRELIMINARY; PRT; 1476 AA.
AC 069387;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
GN GTFB.
OS STREPTOCOCCUS MUTANS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
RN STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MT4239;
RX MEDLINE: 98231643.
RA FUJIWARA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
KIMURA S., HAMADA S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
streptococcus mutans.";
RL FEMS MICROBIOL. LETT. 161:331-336(1998).
DR EMBL: D88657; D1027049; -.
KW TRANSFERASE.
SQ SEQUENCE 1476 AA; 165819 MW; AF6DBA6F CRC32;

Query Match 82.2%; Score 111; DB 2; Length 1476;
Best Local Similarity 94.7%; Pred. No. 2.83e-12;

Matches 18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Db 552 VPYSFIR-AHSEVODLI 569
OY 1 VPYSFIRTAHSEVODLI 19

RESULT 8
ID 069384 PRELIMINARY; PRT; 1476 AA.
AC 069384;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
GN GTFB.
OS STREPTOCOCCUS MUTANS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
RN STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MT4245;
RX MEDLINE: 98231643.
RA FUJIWARA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
KIMURA S., HAMADA S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
streptococcus mutans.";
RL FEMS MICROBIOL. LETT. 161:331-336(1998).
DR EMBL: D88654; D1027045; -.
KW TRANSFERASE.
SQ SEQUENCE 1476 AA; 165792 MW; C76D8483 CRC32;

Query Match 82.2%; Score 111; DB 2; Length 1476;
Best Local Similarity 94.7%; Pred. No. 2.83e-12;

Matches 18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Db 552 VPYSFIR-AHSEVODLI 569
OY 1 VPYSFIRTAHSEVODLI 19

RESULT 9
ID 069381 PRELIMINARY; PRT; 1476 AA.
AC 069381;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
GN GTFB.
OS STREPTOCOCCUS MUTANS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
RN STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MT8148;
RX MEDLINE: 98231643.
RA FUJIWARA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
KIMURA S., HAMADA S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
streptococcus mutans.";
RL FEMS MICROBIOL. LETT. 161:331-336(1998).
DR EMBL: D88651; D1027041; -.
KW TRANSFERASE.
SQ SEQUENCE 1476 AA; 165638 MW; 0868C20A CRC32;

Query Match 82.2%; Score 111; DB 2; Length 1476;
Best Local Similarity 94.7%; Pred. No. 2.83e-12;

Matches 18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Db 552 VPYSFIR-AHSEVODLI 569
OY 1 VPYSFIRTAHSEVODLI 19

RESULT 10
ID 069396 PRELIMINARY; PRT; 1476 AA.
AC 069396;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
GN GTFB.
OS STREPTOCOCCUS MUTANS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
RN STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MT4467;
RX MEDLINE: 98231643.
RA FUJIWARA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
KIMURA S., HAMADA S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
streptococcus mutans.";
RL FEMS MICROBIOL. LETT. 161:331-336(1998).
DR EMBL: D89977; D1027059; -.
KW TRANSFERASE.
SQ SEQUENCE 1476 AA; 165647 MW; 258D4ADD CRC32;

Query Match 82.2%; Score 111; DB 2; Length 1476;
Best Local Similarity 94.7%; Pred. No. 2.83e-12;

Matches 18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Db 552 VPYSFIR-AHSEVODLI 569
OY 1 VPYSFIRTAHSEVODLI 19

RESULT 11
ID 055263 PRELIMINARY; PRT; 1590 AA.
AC 055263;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE GLUCOSYLTRANSFERASE GTF-1.
 OS STREPTOCOCCUS SOBRINUS.
 OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
 OC STREPTOCOCCUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 33478;
 RA SATO S.;
 RL ANN. KAGOSHIMA UNIV. DENTAL SCHOOL 16:23-29(1996).
 DR EMBL: D63570; D1010438;
 DR PFAM: PF00128; alpha-amylase; 1.
 KM TRANSFERASE.
 SO SEQUENCE 1590 AA; 176057 MW; 4718666A CRC32;
 Query Match 71.9%; Score 97; DB 2; Length 1590;
 Best Local Similarity 84.2%; Pred. No. 1.14e-08;
 Matches 16; Conservative 1; Mismatches 1; Indels 1; Gaps 1;
 548 VPYSFIR-AHDSVODLI 565
 :|||||:|||||:|
 QY 1 VPYSFIR-AHDSVODLI 19

RESULT 12
 ID Q59983 PRELIMINARY; PRT; 1590 AA.
 AC Q59983;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-JUN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
 DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (DEXTRANSUCRASE) (SUCROSE 6-GLUCOSYLTRANSFERASE).
 GN GTF1.
 OS STREPTOCOCCUS SOBRINUS.
 OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
 OC STREPTOCOCCUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OM2176;
 RX MEDLINE: 94146405.
 RA SATO S., INOUE M., HANDA N., AIZAWA Y., ISOBE Y., KATAYAMA T.;
 RT "DNA sequence of the glucosyltransferase gene of serotype d Streptococcus sobrinus."
 RT DNA SEQ. 4:19-27(1993).
 CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -
 CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).
 DR EMBL: D13858; D1003482;
 DR PFAM: PF00128; alpha-amylase; 1.
 DE SIGNAL; TRANSFERASE; GLUCOSYLTRANSFERASE.
 GN SIGNAL 1
 GN CHAIN 39 1590 POTENTIAL.
 FT SEQUENCE 1590 AA; 175955 MW; F66A57D7 CRC32;
 SQ
 Query Match 71.9%; Score 97; DB 2; Length 1590;
 Best Local Similarity 84.2%; Pred. No. 1.14e-08;
 Matches 16; Conservative 1; Mismatches 1; Indels 1; Gaps 1;
 Db 548 VPYSFIR-AHDSVODLI 565
 :|||||:|||||:|
 QY 1 VPYSFIR-AHDSVODLI 19

RESULT 13
 ID Q52224 PRELIMINARY; PRT; 1508 AA.
 AC Q52224;
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
 DE GLUCOSYLTRANSFERASE (EC 2.4.1.5) (DEXTRANSUCRASE)
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).
 GN DSRB.
 OS LEUCONOSTOC MESPENTEROIDES.
 OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; LACTOBACILLACEAE;
 OC LEUCONOSTOC.

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NRRL B-1299;
 RA MONCHOIS V., REMAUD-SIMEON M., MONSAN P., WILLENOT R.M.;
 RL FEMS MICROBIOL. LETT. 0:0-0(1998).
 CC -1- CATALYTIC ACTIVITY: SUCROSE +
 CC (1,6-ALPHA-D-GLUCOSYL)(N) - D-FRUCTOSE +
 CC EMBL: AF030129; G2766612;
 DR TRANSFERASE; GLUCOSYLTRANSFERASE.
 KM TRANSFERASE.
 SO SEQUENCE 1508 AA; 168511 MW; 7D3B6FFA CRC32;
 Query Match 69.6%; Score 94; DB 2; Length 1508;
 Best Local Similarity 70.0%; Pred. No. 6.38e-08;
 Matches 14; Conservative 4; Mismatches 1; Indels 1; Gaps 1;
 Db 634 IPNYSFVR-AHDSVOTVIA 652
 :|||||:|||||:|
 QY 1 VPYSFIR-AHDSVODLI 20

RESULT 14
 ID Q48756 PRELIMINARY; PRT; 1290 AA.
 AC Q48756;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE DEXTRANSUCRASE.
 OS LEUCONOSTOC MESPENTEROIDES.
 OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; LACTOBACILLACEAE;
 OC LEUCONOSTOC.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NRRL B1299;
 RX MEDLINE: 97136686.
 RA MONCHOIS V., WILLENOT R.M., REMAUD-SIMEON M., CROUX C., MONSAN P.;
 RT "Cloning and sequencing of a gene coding for a novel dextranucrase from Leuconostoc mesenteroides NRRL B-1299 synthesizing only alpha (1-6) and alpha (1-3) linkages."
 RT GENE 182:23-32(1996).
 RL EMBL: U38181; G1022963;
 DR PFAM: PF00128; alpha-amylase; 1.
 DR SEQUENCE 1290 AA; 145590 MW; F4460B13 CRC32;
 SQ
 Query Match 78.1%; Score 92; DB 2; Length 1290;
 Best Local Similarity 78.9%; Pred. No. 1.99e-07;
 Matches 15; Conservative 2; Mismatches 1; Indels 1; Gaps 1;
 Db 388 PNTSFIR-AHDSVOTIIA 405
 :|||||:|||||:|
 QY 2 PNTSFIR-AHDSVOTIIA 20

RESULT 15
 ID Q55264 PRELIMINARY; PRT; 1449 AA.
 AC Q55264;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE GLUCOSYLTRANSFERASE PRECURSOR.
 GN GTF1.
 OS STREPTOCOCCUS SALIVARIUS.
 OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
 OC STREPTOCOCCUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 95122197.
 RA SIMPSON C.L., GIFFARD P.M., JACQUES N.A.;
 RT "Streptococcus salivarius ATCC 25975 possesses at least two genes coding for primer-independent glucosyltransferases."
 RT INFECT. IMMUN. 63:609-621(1995).
 DR EMBL: U35495; G662379;
 DR PFAM: PF00128; alpha-amylase; 1.

KM SIGNAL: TRANSFERASE. 1 35
FT SIGNAL 1 35
FT CHAIN 36 1449
SQ SEQUENCE 1449 AA: 159984 MW: E9A4BA87 CRC32:

Query Match 60.0%; Score 81; DB 2; Length 1449;
Best Local Similarity 55.0%; Pred. No. 8.67e-05;
Matches 11; Conservative 7; Mismatches 1; Indels 1; Gaps 1;

Db 607 MANYAFVR-AHDSEVOSIG 625
QY 1 VPSYSFIRTAHDSVQDLIA 20

Search completed: Tue Jan 11 15:45:45 2000
Job time : 47 secs.

 WISE (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Msrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Jan 11 15:47:50 2000; Maspar time 2.45 Seconds
 254.053 Million cell updates/sec

Target output not generated.

File: >US-09-290-049-3

Description: (1-22) from US9290049.pep

Sequence: 155
 1 TGARTINGQLLYFRANGVQVG 22

Scoring table: PAM 150
 Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: swiss-prot37
 1:swissprot

Statistics: Mean 29.681; Variance 38.978; scale 0.761

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	155	100.0	1475	1	GLUCOSYLTRANSFERASE-I	9.19e-23
2	114	73.5	1597	1	GLUCOSYLTRANSFERASE-I	4.74e-12
3	112	72.3	1592	1	GLUCOSYLTRANSFERASE-I	1.49e-11
4	106	68.4	1375	1	GLUCOSYLTRANSFERASE-SI	4.46e-10
5	104	67.1	1365	1	GLUCOSYLTRANSFERASE-SI	1.36e-09
6	93	60.0	1430	1	GLUCOSYLTRANSFERASE-S	5.53e-07
7	79	51.0	382	1	GLUCOSYLTRANSFERASE-S	7.50e-04
8	72	46.5	126	1	DEATH ON CURING PROTEIN	2.20e-02
9	70	45.2	529	1	HYPOHETICAL 59.0 KD P	5.58e-02
10	67	43.2	2710	1	TOXIN A	2.19e-01
11	65	41.9	349	1	PHOSPHOTRIESTERASE REL	5.32e-01
12	65	41.9	1396	1	L-SHAPED TAIL FIBRE PR	1.27e+00
13	63	40.6	301	1	ASIALOGALYCOPROTEIN REC	1.27e+00
14	63	40.6	310	1	HYPOHETICAL SUGAR KIN	1.27e+00
15	63	40.6	349	1	PHOSPHOTRIESTERASE REL	1.27e+00
16	63	40.6	396	1	DEOXYRIBIDINE 5'-TRIPHOS	1.27e+00
17	63	40.6	404	1	ALPHA-N-ACETYLDIALACTOS	1.27e+00
18	63	40.6	475	1	CALPHRIN COAT ASSEMBLY	1.27e+00
19	63	40.6	1082	1	RNA-DIRECTED RNA POLYM	1.95e+00
20	62	40.0	464	1	MINOR CAPSID PROTEIN L	1.95e+00
21	62	40.0	467	1	SEXUAL DIFFERENTIATION	1.95e+00
22	62	40.0	524	1	MINOR CAPSID PROTEIN L	1.95e+00
23	61	39.4	336	1	INTEGRASE	2.97e+00

ALIGNMENTS

RESULT ID	1	STANDARD	PRT: 1475 AA.
AC	01-NOV-1988 (REL. 09, CREATED)		
DT	01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)		
DT	15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)		
DE	GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)		
DE	(SDROSE 6-GLUCOSYLTRANSFERASE).		
GN	GTF.		
OS	STREPTOCOCCUS MUTANS.		
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;		
OC	STREPTOCOCCUS.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-GS-5;		
RX	MEDLINE: 87306013.		
RA	SHIMOZA T., UEDA S., KURAMITSU H.K.;		
RT	"Sequence analysis of the gtf gene from Streptococcus mutans.";		
RL	J. BACTERIOL. 169:4263-4270(1987).		
CC	- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.		
CC	- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).		
CC	- SUBCELLULAR LOCATION: SECRETED.		
CC	- DISEASE: DENTAL CARIES.		
CC	- GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF GLUCANS.		
CC	- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-BINDING PROTEIN FROM S. MUTANS.		
CC	*****		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).		
CC	*****		
DR	EMBL: M17361; G153640; -		
DR	PIR: B33135; B33135.		
DR	PIR: PF00128; alpha-amylose; 1.		
DR	TRANSFERASE; GLUCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.		
FT	SIGNAL 1 34		

Query Match 100.0%; Score 155; DB 1; Length 1475;
 Best Local Similarity 100.0%; Pred. No. 9,19e-23;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1300 TGAATNGOLLYFRANGOVYK 1321
 1 TGAATNGOLLYFRANGOVYK 22

RESULT 2
 ID GTF2_STRDO STANDARD; PRT: 1597 AA.
 AC P11001;
 DT 01-JUL-1989 (REL. 11, CREATED)
 DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE GLUCOSYLTRANSFERASE-1 PRECURSOR (EC 2.4.1.5) (GTF-1) (DEXTRANSUCRASE)
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).
 GN GTF1
 OS STREPTOCOCCUS DORNI (STREPTOCOCCUS SOBRIUS).
 OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
 OC STREPTOCOCCUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ME28;
 RX MEDLINE: 87308014.
 RA FERRETTI J.J., GILPIN M.L., RUSSELL R.R.B.;
 RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus
 sobrius ME28.";
 RL J. BACTERIOL. 169:4271-4278(1987).
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
 OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDATE THE
 AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
 CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -
 D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- DISEASE: DENTAL CARIES.
 CC -1- GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED
 GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE
 GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF
 GLUCANS.
 CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
 BINDING PROTEIN FROM S.MUTANS.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to [license@isb-sid.ch](mailto:license@isb-sib.ch)).
 CC
 CC EMBL: M17391; G153647; -
 DR PIR: P00128; alpha-amylase; 1.
 DR PIR: P00128; alpha-amylase; 1.
 KW TRANSFERASE; GLYCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.
 FT SIGNAL 1 38
 FT CHAIN 39 1597
 FT DOMAIN 39 1050
 FT DOMAIN 1099 1597
 FT DOMAIN 1099 1597
 FT REPEAT 1099 1132
 FT A REPEAT.
 FT 5 X TANDEM REPEATS.

Query Match 73.5%; Score 114; DB 1; Length 1597;
 Best Local Similarity 77.3%; Pred. No. 4,74e-12;
 Matches 17; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 1303 TGAATNGOLLYFRANGOVYK 1324
 1 TGAATNGOLLYFRANGOVYK 22

Query Match 73.5%; Score 114; DB 1; Length 1597;
 Best Local Similarity 77.3%; Pred. No. 4,74e-12;
 Matches 17; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 1303 TGAATNGOLLYFRANGOVYK 1324
 1 TGAATNGOLLYFRANGOVYK 22

RESULT 3
 ID GTF2_STRDO STANDARD; PRT: 1592 AA.
 AC P27470;
 DT 01-AUG-1992 (REL. 23, CREATED)
 DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE GLUCOSYLTRANSFERASE-1 PRECURSOR (EC 2.4.1.5) (GTF-1) (DEXTRANSUCRASE)
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).
 GN STREPTOCOCCUS DORNI (STREPTOCOCCUS SOBRIUS).
 OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
 OC STREPTOCOCCUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-6715;
 RX MEDLINE: 91123227.
 RA ABO H., MATSUMURA T., KODAMA T., OHTA H., FUKUI K., KATO K.,
 RA KAGAWA H.;
 RT "Peptide sequences for sucrose splitting and glucan binding within
 Streptococcus sobrius glucosyltransferase (water-insoluble glucan
 synthetase).";
 RL J. BACTERIOL. 173:989-996(1991).
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
 OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDATE THE
 AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
 CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -
 D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- DISEASE: DENTAL CARIES.
 CC -1- GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED
 GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE
 GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF
 GLUCANS.
 CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
 BINDING PROTEIN FROM S.MUTANS.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to [license@isb-sid.ch](mailto:license@isb-sib.ch)).
 CC
 CC EMBL: D90213; G217033; -
 DR PIR: A38175; A38175.
 DR PIR: A38175; A38175.
 KW TRANSFERASE; GLYCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.
 FT SIGNAL 1 38
 FT CHAIN 39 1592
 FT DOMAIN 39 1044
 FT DOMAIN 1093 1592
 FT DOMAIN 1093 1592
 FT REPEAT 1093 1592
 FT 6.5 X TANDEM REPEATS.

```

FT REPEAT 1093 1142 1.
FT REPEAT 1158 1207 2.
FT REPEAT 1222 1272 3.
FT REPEAT 1287 1337 4.
FT REPEAT 1402 1451 5.
FT REPEAT 1514 1563 6.
FT REPEAT 1577 1592 7. (INCOMPLETE).
SO SEQUENCE 1592 AA; 176167 MW; 2734237A CRC32;

Query Match 72.3%; Score 112; DB 1; Length 1592;
Best Local Similarity 77.3%; Pred. No. 1,49e-11;
Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 1298 TGAORINGOKLYFRANGOOVKG 1319
Oy 1 TGAORINGOKLYFRANGVOVKG 22
||||| ||||| |||||

RESULT 4
ID GTFC STRMG STANDARD: PRT: 1375 AA.
AC 01-NOV-1988 (REL. 09, CREATED)
T 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
D 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GLUCOSYLTRANSFERASE-SI PRECURSOR (EC 2.4.1.5) (GTF-SI)
DE (DEXTRANSUCRASE) (SUCROSE 6-GLUCOSYLTRANSFERASE).
GN GTFC.
OS STREPTOCOCCUS MUTANS.
OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-GS-5;
RX MEDLINE: 89137980.
RA UEDA S., SHIROZA T., KURAMITSU H.K.;
RT "Sequence analysis of the gtfc gene from Streptococcus mutans GS-5.";
RL GENE 69:101-109(1988).
[2]
RP SEQUENCE OF 1-349 FROM N.A.
RC STRAIN-GS-5;
RX MEDLINE: 87308013.
RA SHIROZA T., UEDA S., KURAMITSU H.K.;
RL "Sequence analysis of the gtfr gene from Streptococcus mutans.";
RT J. BACTERIOL. 169:4263-4270(1987).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) =
CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DISEASE: DENTAL CARIES.
CC -1- GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED
CC GLUCOSE AND SOME 1,6 LINKAGES). GTF-S SYNTHESIZES WATER-SOLUBLE
CC GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF
CC GLUCANS.
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S. MUTANS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: M2054; G153643; -
DR EMBL: M17361; G153641; -
DR PIR: J70345; J70345.
DR PIR: C3135; C3135.
DR PFAM: PF00128; alpha-amylase; 1.
DR TRANSFAM: GLYCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.

```

```

FT SIGNAL 1 34 GLUCOSYLTRANSFERASE-ST.
FT CHAIN 35 1375 CATALYTIC (APPROXIMATE).
FT DOMAIN 35 1050
FT DOMAIN 1126 1375 GLUCAN-BINDING (APPROXIMATE).
FT DOMAIN 1126 1375 2.4 A, 1 C AND 1 AC REPEATS.
FT REPEAT 1126 1159 A REPEAT.
FT REPEAT 1169 1200 A REPEAT.
FT REPEAT 1227 1238 C REPEAT.
FT REPEAT 1253 1303 AC REPEAT.
FT REPEAT 1318 1330 A REPEAT (INCOMPLETE).
SQ SEQUENCE 1375 AA; 153022 MW; 3EA3727E CRC32;

Query March 68.4%; Score 106; DB 1; Length 1375;
Best Local Similarity 68.2%; Pred. No. 4,46e-10;
Matches 15; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 1264 TGTVTNGCARLYFKFNGVQANG 1285
Cy 1 TGAATNGQLLYFRANGVQYKNG 22
||:|:|||||:|||||:|
|:|:|:|:|:|:|:|:|:|

RESULT 5
ID GTF5_STRDO STANDARD; PRT; 1365 AA.
AC P29336;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).
GN GTF5.
OS STREPTOCOCCUS DOWNII (STREPTOCOCCUS SOBRIINUS).
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MFE28:
RX MEDLINE; 90316665.
RA GILMORE K.S., RUSSELL R.R., FERRETTI J.J.;
RT "Analysis of the Streptococcus downii gtf gene, which specifies a
RT glucosyltransferase that synthesizes soluble glucans.";
RL INFECT. IMMUN. 58:2452-2458(1990).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =
CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).
CC -1- ENZYME REGULATION: GLUCAN SYNTHESIS BY GTF-S IS INDEPENDENT OF
CC PRIMER GLUCAN OMLIKE GTF-I.
CC -1- DISEASE: DENTAL CARIES.
CC -1- GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE).
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S. MUTANS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M30943; G153653; -.
DR PIR; A41483; A41483.
DR PFAM; PF00128; alpha-amylase; 1.
KM TRANSFERASE; GLYCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.
KM SIGNAL 1 36 OR 37 (POTENTIAL).
FT CHAIN 37 1365 GLUCOSYLTRANSFERASE-S.
FT DOMAIN 37 1050 CATALYTIC (APPROXIMATE).
FT DOMAIN 1083 1365 GLUCAN-BINDING (APPROXIMATE).
FT DOMAIN 1083 1365 4.5 X TANDEM REPEATS.
FT REPEAT 1083 1131 1.
FT REPEAT 1150 1199 2.

```

FT REPEAT 1225 1274 3.
 FT REPEAT 1289 1339 4.
 FT REPEAT 1353 1365 5 (INCOMPLETE).
 SQ SEQUENCE 1365 AA; 151590 MM; 156F8B31 CRC32;

Query Match 67.1% Score 104; DB 1; Length 1365;
 Best Local Similarity 63.6% Pred. No. 1.35e-09;
 Matches 14; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

DB 1235 TGEOTIDGKVFQDNGVQVG 1256
 1 TGAATNGQLLYFRANGVQVG 22

RESULT 6
 ID GTFD-STRMU STANDARD; PRT; 1430 AA.
 AC P49331;
 DT 01-FEB-1996 (REL. 33, CREATED)
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
 DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE GLUCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)
 DE (SUCCOSE 6-GLUCOSYLTRANSFERASE).
 GN GTFD.
 OS STREPTOCOCCUS MUTANS.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
 OC STREPTOCOCCUS.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-GS-5;
 RX MEDLINE: 91100958.
 RA HONDA O., KATO C., KURAMITSU H.K.;
 RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding
 the glucosyltransferase-S enzyme."
 RL J. GEN. MICROBIOL. 136:2099-2105(1990).
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
 OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
 AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
 CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -
 D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- DISEASE: DENTAL CARIES.
 CC -1- GTF-S SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED
 GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE
 GLUCANS (ALPHA 1,6-GLUCOSE). GTF-S SYNTHESIZES BOTH FORMS OF
 GLUCANS.
 CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
 BINDING PROTEIN FROM S. MUTANS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M29296; G153645; -
 CC PFAM: PF00128; alpha-amylase: 1.
 CC TRANSFERASE: GLUCOSYLTRANSFERASE: SIGNAL; REPEAT: DENTAL CARIES.
 CC SIGNAL 1
 FT CHAIN 1 1430
 FT DOMAIN 1232 1423 GLUCOSYLTRANSFERASE-S.
 FT REPEAT 1232 1295 3 X 63 AA APPROXIMATE TANDEM REPEATS.
 FT REPEAT 1296 1359 2.
 FT REPEAT 1360 1423 3.
 SO SEQUENCE 1430 AA; 159765 MM; D3DE3681 CRC32;

OY 1 TGAATNGQLLYFRANGVQVG 22

RESULT 7
 ID YJ07 YEAST STANDARD; PRT; 382 AA.
 AC P47007;
 DT 01-FEB-1996 (REL. 33, CREATED)
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
 DE 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 44.9 KD PROTEIN IN INO1-1DS2 INTERGENIC REGION.
 GN YJ147C OR J0639.
 OS SACCAROMYCES CEREVISIAE (BAKER'S YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCAROMYCETALES;
 OC SACCAROMYCETACEAE; SACCAROMYCES.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / FY1679;
 RX MEDLINE: 96408771.
 RA KATSIOLOU C., TZEREMIA M., TAVERNARAKIS N., ALEXANDRAKI D.;
 RT "Sequence analysis of a 40.7 kb segment from the left arm of yeast
 chromosome X reveals 14 known genes and 13 new open reading frames
 including homologues of genes clustered on the right arm of
 chromosome XI."
 RL YEAST 12:787-797(1996).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z49422; G1015561; -
 CC EMBL: X67371; G854550; -
 CC HYPOTHETICAL PROTEIN.
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 382 AA; 44862 MM; 5B128AB9 CRC32;

Query Match 51.0% Score 79; DB 1; Length 382;
 Best Local Similarity 38.9% Pred. No. 7.50e-04;
 Matches 7; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

DB 170 RYVISTILYRTHDVQIR 187
 4 RINGQLLYFRANGVQVG 21

RESULT 8
 ID DOC_BPPI STANDARD; PRT; 126 AA.
 AC Q06259;
 DT 01-FEB-1995 (REL. 31, CREATED)
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
 DE DEATH ON CURING PROTEIN.
 GN DOC.
 OS BACTERIOPHAGE P1.
 OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; MYOVIRIDAE.
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94016561.
 RA LEHNHERR H., MAGUIN E., JAFRI S., YARMOLINSKY M.B.;
 RT "Plasmid addiction genes of bacteriophage P1: doc, which causes cell
 death on curing of prophage, and phd, which prevents host death when
 prophage is retained."
 RL J. MOL. BIOL. 233:414-428(1993).
 CC -1- FUNCTION: KILLS CELLS. DOC AND PHD PROTEIN FUNCTION IN UNISSON TO
 STABILIZE PLASMID NUMBER BY INDUCING A LETHAL RESPONSE TO PLASMID
 LOSS.
 CC -1- THE CONCENTRATION OF PHD IN P1 LYSOGENS IS FAR GREATER THAN THAT
 OF THE POISON IT ANTAGONIZES. SUCH AN EXCESS MAY ASSUME THE
 WELL-BEING OF CARRIERS OF THE ADDICTING PLASMID.
 CC
 CC

CC BACTERIA/FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VPI 10463;
RX MEDLINE: 90221894.
RA SADEBORN M., VON EICHEL-STREIBER C.;
RT "Nucleotide sequence of Clostridium difficile toxin A";
RL NCLEOT ACIDS RES. 18:1629-1630(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-VPI 10463;
RX MEDLINE: 90129305.
RA DOVE C.H., WANG S.Z., PRICE S.B., PHELPS C.J., LYERLY D.M.,
RA WILKINS T.W., JOHNSON J.L.;
RT "Molecular characterization of the Clostridium difficile toxin A
gene."
RL INFECT. IMMUN. 58:480-488(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-VPI 10463;
RA VON EICHEL-STREIBER C.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DDJ DATA BANKS.
CC -I DOMAIN: THE C-TERMINAL PART OF TOXIN A CONSISTS OF A 833 AA
CC REPETITIVE STRUCTURE. THIS PART OF TOXIN A IS COMPOSED OF FIVE
CC DIFFERENT OLIGOPEPTIDES.
CC -I DISEASE: CLOSTRIDIUM DIFFICILE PRODUCES TWO TOXINS, AN
CC ENTEROTOXIN CALLED A AND CYTOTOXIN B. ONLY AFTER THE ENTERAL
CC DELIVERY OF THE ENTEROTOXIN A MAY THE CHARACTERISTIC DISEASE
CC CALLED PSEUDOMEMBRANOUS COLITIS BE INDUCED.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X51797; G40441; -;
DR EMBL: M30307; G144926; -;
DR EMBL: X92982; E212010; -;
DR PIR: S08638; S08638.
KW TOXIN; ENTEROTOXIN.
SQ SEQUENCE 2710 AA; 308052 MW; D04C4C08 CRC32;

Query March 43.2% Score 67; DB 1; Length 2710;
Best Local Similarity 52.2%; Pred. No. 2.19e-01;
Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

Db 2460 TGLRTIDGKKYYNTNAVAVTG 2482
|||::| | : ||
QY 1 TGARTINGQLTFRAN-GVGYSKG 22

RESULT 11
ID PTER_MOUSE STANDARD; PRT; 349 AA.
AC 060866;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE PHOSPHOTRIESTERASE RELATED PROTEIN (PARATHION HYDROLASE-RELATED
DE PROTEIN).
DE PTER.
GN MUS MUSCULUS (MOUSE).
OS EURARCTICA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUETHERIA;
OC RODENTIA; SCIURGINATHI; MORIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6J; TISSUE-KIDNEY;
RX MEDLINE: 96194889
RA HOU X., MASER R.L., MAGENHEIMER B.S., CALVERT J.P.;
RT "A mouse kidney- and liver-expressed cDNA having homology with a

RT prokaryotic parathion hydrolase (phosphotriesterase)-encoding gene:
 RT abnormal expression in injured and polycystic kidneys.
 RL GENE 168:157-163(1996).
 CC -1- COFACTOR: CONTAINS 2 MOLES OF ZINC PER SUBUNIT (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN KIDNEY AND LIVER.
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOTRIESTERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U28016; G881499; -
 DR MGD: MGI:107372; PTER.
 KM HYDROLASE; ZINC.
 CC METAL 26 26 ZINC 1 (BY SIMILARITY).
 CC METAL 28 28 ZINC 1 (BY SIMILARITY).
 CC METAL 169 169 ZINC 1 AND 2 (BY SIMILARITY).
 CC METAL 201 201 ZINC 2 (BY SIMILARITY).
 CC METAL 230 230 ZINC 2 (BY SIMILARITY).
 CC METAL 298 298 ZINC 1 (BY SIMILARITY).
 CC SEQUENCE 349 AA; 39218 MW; 58404EEP CRC32.
 SQ
 Query Match 41.9%; Score 65; DB 1; Length 349;
 Best Local Similarity 53.8%; Pred. No. 5.32e-01;
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 DB 79 AIRBELLYFKAKG 91
 QY 1:|||||:1
 5 TINGOLLYFRANG 17
 RESULT 12
 ID VITE.BPTS STANDARD; PRT; 1396 AA.
 AC P13390; Q48502;
 DT 01-JAN-1990 (REL. 13, CREATED)
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE L-SHAPED TAIL FIBRE PROTEIN (LTF PROTEIN).
 GN LTF
 OS BACTERIOPHAGE T5.
 OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; TAILED PHAGES; SIPHOVIRIDAE.
 RN 11
 RP SEQUENCE FROM N.A.
 RM MEDLINE: 95309401.
 RA KALIMAN A.V., KULSHIN V.E., SHLYAPNIKOV M.G., KSENZENKO V.N.,
 RA KRYUKOV V.M.;
 RT The nucleotide sequence of the bacteriophage T5 ltf gene.*;
 RL FEBS LETT. 366:46-48(1995).
 CC [2]
 CC SEQUENCE FROM N.A.
 RP KALIMAN A.V.;
 RL SUBMITTED (JAN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
 RM MEDLINE: 88289370.
 RA KALIMAN A.V., KRYUKOV V.M., BAYEV A.A.;
 RT The nucleotide sequence of bacteriophage T5 DNA at the region
 RT between early and late genes.*;
 RL NUCLEIC ACIDS RES. 16:6230-6230(1988).
 CC -1- FUNCTION: NONSENSE-TAIL PROTEIN THAT MEDIATES BINDING TO THE
 CC POLYMANNOSE O ANTIGEN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC EMBL: X69460; G15416; -
 DR EMBL: AJ001191; E1237676; -
 DR PIR: S01982; S01982.
 KM LATE PROTEIN.
 FT CONFLICT 986 986 V -> A (IN REF. 2).
 SQ SEQUENCE 1396 AA; 147989 MW; 78D7BBAC CRC32;
 V
 Query Match 41.9%; Score 65; DB 1; Length 1396;
 Best Local Similarity 70.0%; Pred. No. 5.32e-01;
 Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 DB 922 GQMTTPRNG 931
 QY 1:|||||:1
 8 GOLLYFRANG 17
 RESULT 13
 ID LECT1.MOUSE STANDARD; PRT; 301 AA.
 AC P24721;
 DT 01-MAR-1992 (REL. 21, CREATED)
 DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
 DE ASIALOGLYCOPROTEIN RECEPTOR 2 (HEPATIC LECTIN 2) (HNL-2) (ASGPR).
 GN ASGR2 OR ASGR-2.
 OS MUS MUSCULUS (MOUSE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC RODENTIA; SCIROGNATHI; MURIDAE; MURINAE; MUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C57BL/6; TISSUE-LIVER;
 RM MEDLINE: 91027942.
 RA SANFORD J.P., DOYLE D.;
 RT Mouse asialoglycoprotein receptor cDNA sequence: conservation of
 RT receptor genes during mammalian evolution.*;
 RL BLOCHIM. BIOPHYS. ACTA 1087:259-261(1990).
 CC -1- FUNCTION: MEDIATES THE ENDOCYTOSIS OF PLASMA GLYCOPROTEINS TO
 CC WHICH THE TERMINAL SIALIC ACID RESIDE ON THEIR COMPLEX
 CC CARBOHYDRATE MOIETIES HAS BEEN REMOVED. THE RECEPTOR RECOGNIZES
 CC TERMINAL GALACTOSE AND N-ACETYLGLUCOSAMINE UNITS. AFTER LIGAND
 CC BINDING TO THE RECEPTOR, THE RESULTING COMPLEX IS INTERNALIZED AND
 CC TRANSPORTED TO A SORTING ORGANELLE, WHERE RECEPTOR AND LIGAND ARE
 CC DISASSOCIATED. THE RECEPTOR THEN RETURNS TO THE CELL MEMBRANE
 CC SURFACE.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN HEPATIC PARENCHYMAL
 CC CELLS.
 CC -1- CALCIUM IS REQUIRED FOR LIGAND BINDING.
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X53042; G53105; -
 DR PIR: S13165; S13165.
 DR MGD: MGI:88082; ASGR2.
 DR PROSITE: PS00615; C_TYPE_LLECTIN_1; 1.
 DR PROSITE: PS00615; C_TYPE_LLECTIN_2; 1.
 DR PFAM: PF00059; lectin_c; 1.
 DR HSSP: P06734; 1KIE.
 KM LECTIN; GLYCOPROTEIN RECEPTOR; ENDOCYTOSIS; TRANSMEMBRANE;
 KM CALCIN; SIGNAL-ANCHOR; PHOSPHORYLATION.
 FT DOMAIN 1 58 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 59 79 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
 FT DOMAIN 80 301 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 169 295 C-TYPE LECTIN (LONG FORM).
 FT DISULFID 170 181 BY SIMILARITY.
 FT DISULFID 198 293 BY SIMILARITY.

FT DISULFID 271 285 BY SIMILARITY.
 FT CARBOHYD 97 POTENTIAL.
 FT CARBOHYD 165 165 POTENTIAL.
 FT CARBOHYD 298 298 POTENTIAL.
 SQ SEQUENCE 301 AA: 34907 MW: 3559D5C7 CRC32:

Query Match 40.68; Score 63; DB 1; Length 301;
 Best Local Similarity 50.08; Pred. No. 1.27e+00;
 Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 153 RUTCOLAYFQSNQTE 168
 11: 1111111111
 QY 4 RTINGOLLYFRANGVQ 19

RESULT 14
 ID YC08-PYRHO STANDARD; PRT; 310 AA.
 AC 059128;
 DT 15-DEC-1998 (REL. 37, CREATED)
 DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DT HYPOTHETICAL SUGAR KINASE PHCC008.
 DT PHCC008.
 OS PYROCOCUS HORIKOSHII
 OC ARCHAEA; EURYARCHAEOTA; THERMOCOCCALES; THERMOCOCCACEAE; PYROCOCUS.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-OT3;
 RX MEDLINE: 98344137.
 RA KAMARABYASTI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
 RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOTAMA A., NAGAI Y.,
 RA SAKAI M., OGURA K., OTSUKA R., NAKAZAWA H., TAKAMITA M., OHFURU Y.,
 RA FUNAHASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
 RA AOKI K., KIKUCHI H.;
 RT "Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, Pyrococcus horikoshii OT3";
 RL DNA RES. 5:55-76(1998).
 CC -1- SIMILARITY: BELONGS TO THE PFKB FAMILY OF CARBOHYDRATE KINASES.
 CC -1- SIMILARITY: BELONGS TO THE PFKB FAMILY OF CARBOHYDRATE KINASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AB009517; D1028566; -
 DR PROSITE: PS00583; PFKB_KINASES_1; 1.
 DR PROSITE: PS00584; PFKB_KINASES_2; 1.
 DR HYPOTHETICAL PROTEIN; TRANSFERASE; KINASE.
 SQ SEQUENCE 310 AA: 33986 MW: B884F991 CRC32:

Query Match 40.68; Score 63; DB 1; Length 310;
 Best Local Similarity 36.88; Pred. No. 1.27e+00;
 Matches 7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

Db 194 KASEEVLLENQVEYKG 212
 11: 1111111111
 QY 4 RTINGOLLYFRANGVQ 22

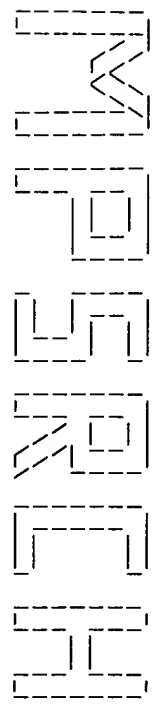
RESULT 15
 ID PTER-RAT STANDARD; PRT; 349 AA.
 AC 063530;
 DT 15-DEC-1998 (REL. 37, CREATED)
 DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE PHOSPHOTRIESTERASE RELATED PROTEIN (PARATHION HYDROLASE-RELATED PROTEIN) (RESINIFEROXIN-BINDING PHOSPHOTRIESTERASE-RELATED PROTEIN).
 DE PTER OR RPR-1.
 OS RATUS NORVEGICUS (RAT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATIUS.
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 9739344.
 RA DAVIES J.A., BUCHANAN V., KRYLOVA O., NINKINA N.N.;
 RT "Molecular cloning and expression pattern of rpr-1, a resiniferatoxin-binding, phosphotriesterase-related protein, expressed in rat kidney tubules.";
 RL FEBS LETT. 410:378-382(1997).
 CC -1- FUNCTION: BINDS RESINIFEROXIN, A VANILLOID THAT DESSENSITIZES NOCICEPTIVE NEURONS.
 CC -1- CORFACTOR: CONTAINS 2 MOLES OF ZINC PER SUBUNIT (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: EXPRESSED PRIMARILY IN PROXIMAL TUBULES OF THE KIDNEY.
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOTRIESTERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X99477; E254899; -
 DR KW HYDROLASE; ZINC.
 FT METAL 26 26 ZINC 1 (BY SIMILARITY).
 FT METAL 28 28 ZINC 1 (BY SIMILARITY).
 FT METAL 169 169 ZINC 1 AND 2 (BY SIMILARITY).
 FT METAL 201 201 ZINC 2 (BY SIMILARITY).
 FT METAL 230 230 ZINC 2 (BY SIMILARITY).
 FT METAL 298 298 ZINC 1 (BY SIMILARITY).
 SQ SEQUENCE 349 AA: 38931 MW: A35021FD CRC32:

Query Match 40.68; Score 63; DB 1; Length 349;
 Best Local Similarity 46.28; Pred. No. 1.27e+00;
 Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 79 AVREELLYFRKAG 91
 11: 1111111111
 QY 5 TINGOLLYFRANG 17

Search completed: Tue Jan 11 15:47:57 2000
 Job time : 7 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MSrch_PP protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Jan 11 15:48:14 2000; MasPar time 5.05 Seconds
237.695 Million cell updates/sec

Output not generated.

Title: >US-09-290-049-3

Description: (1-22) from US09290049.pep

Perfect Score: 155

Sequence: 1 TGARTINGQLLYFRANGVQVKG 22

Scoring table: PAM 150
Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: splrembl9

1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mic 8:sp-organalle
9:sp-phase 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 29.020; Variance 39.036; scale 0.743

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	143	92.3	1476	2	069384	GLUCOSYLTRANSFERASE-I.	5.35e-19
2	143	92.3	1476	2	069396	GLUCOSYLTRANSFERASE-I.	5.35e-19
3	143	92.3	1476	2	069381	GLUCOSYLTRANSFERASE-I.	5.35e-19
4	143	92.3	1476	2	069387	GLUCOSYLTRANSFERASE-I.	5.35e-19
5	143	92.3	1476	2	069390	GLUCOSYLTRANSFERASE-I.	5.35e-19
6	143	92.3	1476	2	069388	GLUCOSYLTRANSFERASE-I.	5.35e-19
7	143	92.3	1476	2	069387	GLUCOSYLTRANSFERASE-I.	5.35e-19
8	143	92.3	1476	2	069382	GLUCOSYLTRANSFERASE-I.	5.35e-19
9	143	92.3	1476	2	069381	GLUCOSYLTRANSFERASE-I.	5.35e-19
10	143	92.3	1476	2	069383	GLUCOSYLTRANSFERASE-I.	5.35e-19
11	143	92.3	1476	2	069383	GLUCOSYLTRANSFERASE-I.	5.35e-19
12	143	92.3	1476	2	069383	GLUCOSYLTRANSFERASE-I.	5.35e-19
13	143	92.3	1476	2	069383	GLUCOSYLTRANSFERASE-I.	5.35e-19
14	143	92.3	1476	2	069383	GLUCOSYLTRANSFERASE-I.	5.35e-19
15	143	92.3	1476	2	069383	GLUCOSYLTRANSFERASE-I.	5.35e-19
16	143	92.3	1476	2	069383	GLUCOSYLTRANSFERASE-I.	5.35e-19
17	143	92.3	1476	2	069383	GLUCOSYLTRANSFERASE-I.	5.35e-19
18	143	92.3	1476	2	069383	GLUCOSYLTRANSFERASE-I.	5.35e-19
19	143	92.3	1476	2	069383	GLUCOSYLTRANSFERASE-I.	5.35e-19
20	143	92.3	1476	2	069383	GLUCOSYLTRANSFERASE-I.	5.35e-19

21	93	60.0	1462	2	069392	GLUCOSYLTRANSFERASE-S.	1.21e-06
22	93	60.0	1462	2	069389	GLUCOSYLTRANSFERASE-S.	1.21e-06
23	93	60.0	1462	2	069389	GLUCOSYLTRANSFERASE-S.	1.21e-06
24	91	58.7	1508	2	052224	GLUCOSYLTRANSFERASE (E	3.47e-06
25	88	56.8	1563	2	054447	GLUCOSYLTRANSFERASE	1.64e-05
26	88	56.8	1577	2	055265	GLUCOSYLTRANSFERASE PR	1.64e-05
27	83	53.5	330	2	055228	DEXTRANASE INHIBITOR P	2.08e-04
28	77	49.7	329	1	050120	329AA LONG HYPOTHETICA	3.99e-03
29	72	46.5	126	2	047172	DNA SEQUENCE UPSTREAM	4.28e-02
30	72	46.5	126	2	047173	ANALOGUE.	4.28e-02
31	72	46.5	126	2	047171	ANALOGUE.	4.28e-02
32	71	45.8	425	2	067315	UDP-N-ACETYLGLUCOSAMIN	6.80e-02
33	68	43.9	831	2	050076	CHITINASE B (EC 3.2.1.	2.67e-01
34	67	43.2	245	1	058723	HYPOTHETICAL PROTEIN M	4.18e-01
35	65	41.9	446	11	063237	HEPARIN-BINDING FIBROB	1.01e+00
36	65	41.9	446	11	063236	HEPARIN-BINDING FIBROB	1.01e+00
37	64	41.3	221	5	044353	GATA TRANSCRIPTION FAC	1.56e+00
38	64	41.3	669	10	081752	HYPOTHETICAL 75.2 KD P	1.56e+00
39	64	41.3	829	5	023290	SIMILAR TO SERINE/THRE	1.56e+00
40	63	40.6	240	1	029326	CONSERVED HYPOTHETICAL	2.40e+00
41	63	40.6	371	10	081753	HYPOTHETICAL 42.3 KD P	2.40e+00
42	63	40.6	566	2	P72770	HIGH AFFINITY SULFATE	2.40e+00
43	63	40.6	574	10	081754	HYPOTHETICAL 65.1 KD P	2.40e+00
44	62	40.0	1360	5	094649	AARF2 PROTEIN.	3.68e+00
45	62	40.0	2364	2	046342	CYTOTOXIN L.	3.68e+00

ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	1476 AA.
1	069384			
AC	069384			
DT	01-AUG-1998 (TREMBLREL. 07, CREATED)			
DT	01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	GLUCOSYLTRANSFERASE-I.			
GN	GTFB.			
OS	STREPTOCOCCUS MUTANS.			
OC	BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;			
OC	STREPTOCOCCUS.			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-MT4245.			
RC	MEDLINE: 98231643.			
RA	FUJIMURA T., TERAO Y., HOSHINO T., KAMABATA S., OOSHIMA T., SOBUE S.,			
RA	KIMURA S., HAMADA S.,			
RT	"Molecular analyses of glucosyltransferase genes among strains of			
RT	streptococcus mutans."			
RT	FEMS MICROBIOL. LETT. 161:331-336(1998).			
RL	EMBL: D88654; D1027045; .			
DR	TRANSFERASE.			
KW	SEQUENCE 1476 AA; 165792 MW; C76D8483 CRC32;			

Query Match 92.3%; Score 143; DB 2; Length 1476;
Best Local Similarity 95.5%; Pred. No. 5.35e-19;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB	1301 TGARTINGQLLYFRANGVQVKG 1322
QY	1 TGARTINGQLLYFRANGVQVKG 22
RESULT	2
ID	069396
AC	069396
DT	01-AUG-1998 (TREMBLREL. 07, CREATED)
DT	01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE	GLUCOSYLTRANSFERASE-I.
GN	GTFB.
OS	STREPTOCOCCUS MUTANS.
OC	BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC	STREPTOCOCCUS.

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MT4467;
 RX MEDLINE: 98231643.
 RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
 RA KIMURA S., HAMADA S.;
 RT "Molecular analyses of glucosyltransferase genes among strains of
 RT Streptococcus mutans.";
 RL FEMS MICROBIOL. LETT. 161:331-336(1998).
 DR EMBL: D89977; D1027059; -
 KW TRANSFERASE.
 SO SEQUENCE 1476 AA; 165647 MW; 258DAAD CRC32;

Query Match 92.3%; Score 143; DB 2; Length 1476;
 Best Local Similarity 95.5%; Pred. No. 5.35e-19;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 1301 TGARTINGOLLYFRANGVOYKG 1322
 1 TGARTINGOLLYFRANGVOYKG 22

RESULT 3
 ID 069381 PRELIMINARY; PRT; 1476 AA.
 AC 069381;
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE GLUCOSYLTRANSFERASE-I.
 GN GTFB.
 OS STREPTOCOCCUS MUTANS.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
 OC STREPTOCOCCUS.
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MT8148;
 RX MEDLINE: 98231643.
 RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
 RA KIMURA S., HAMADA S.;
 RT "Molecular analyses of glucosyltransferase genes among strains of
 RT Streptococcus mutans.";
 RL FEMS MICROBIOL. LETT. 161:331-336(1998).
 DR EMBL: D88651; D1027041; -
 KW TRANSFERASE.
 SO SEQUENCE 1476 AA; 165638 MW; 0868C20A CRC32;

Query Match 92.3%; Score 143; DB 2; Length 1476;
 Best Local Similarity 95.5%; Pred. No. 5.35e-19;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 1301 TGARTINGOLLYFRANGVOYKG 1322
 1 TGARTINGOLLYFRANGVOYKG 22

RESULT 4
 ID 069387 PRELIMINARY; PRT; 1476 AA.
 AC 069387;
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE GLUCOSYLTRANSFERASE-I.
 GN GTFB.
 OS STREPTOCOCCUS MUTANS.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
 OC STREPTOCOCCUS.
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MT4239;
 RX MEDLINE: 98231643.
 RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
 RA KIMURA S., HAMADA S.;
 RT "Molecular analyses of glucosyltransferase genes among strains of
 RT Streptococcus mutans.";
 RL FEMS MICROBIOL. LETT. 161:331-336(1998).
 DR EMBL: D88658; D1027050; -
 KW TRANSFERASE.
 SO SEQUENCE 1455 AA; 163046 MW; 7659CFBC CRC32;

RT Streptococcus mutans.";
 RL FEMS MICROBIOL. LETT. 161:331-336(1998).
 DR EMBL: D88657; D1027049; -
 KW TRANSFERASE.
 SO SEQUENCE 1476 AA; 165819 MW; AF6DBA6F CRC32;

Query Match 92.3%; Score 143; DB 2; Length 1476;
 Best Local Similarity 95.5%; Pred. No. 5.35e-19;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 1301 TGARTINGOLLYFRANGVOYKG 1322
 1 TGARTINGOLLYFRANGVOYKG 22

RESULT 5
 ID 069390 PRELIMINARY; PRT; 1476 AA.
 AC 069390;
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE GLUCOSYLTRANSFERASE-I.
 GN GTFB.
 OS STREPTOCOCCUS MUTANS.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
 OC STREPTOCOCCUS.
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MT4251;
 RX MEDLINE: 98231643.
 RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
 RA KIMURA S., HAMADA S.;
 RT "Molecular analyses of glucosyltransferase genes among strains of
 RT Streptococcus mutans.";
 RL FEMS MICROBIOL. LETT. 161:331-336(1998).
 DR EMBL: D88660; D1027053; -
 KW TRANSFERASE.
 SO SEQUENCE 1476 AA; 165833 MW; 9119C72B CRC32;

Query Match 92.3%; Score 143; DB 2; Length 1476;
 Best Local Similarity 95.5%; Pred. No. 5.35e-19;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 1301 TGARTINGOLLYFRANGVOYKG 1322
 1 TGARTINGOLLYFRANGVOYKG 22

RESULT 6
 ID 069388 PRELIMINARY; PRT; 1455 AA.
 AC 069388;
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE GLUCOSYLTRANSFERASE-SI.
 GN GTFB.
 OS STREPTOCOCCUS MUTANS.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
 OC STREPTOCOCCUS.
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MT4239;
 RX MEDLINE: 98231643.
 RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
 RA KIMURA S., HAMADA S.;
 RT "Molecular analyses of glucosyltransferase genes among strains of
 RT Streptococcus mutans.";
 RL FEMS MICROBIOL. LETT. 161:331-336(1998).
 DR EMBL: D88658; D1027050; -
 KW TRANSFERASE.
 SO SEQUENCE 1455 AA; 163046 MW; 7659CFBC CRC32;

Query Match 78.1%; Score 121; DB 2; Length 1455;

QY 1 TGARTINGQLLYFRANGVQVKG 22

AC 000599;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE GLUCOSYLTRANSFERASE S PRECURSOR (EC 2.4.1.5) (GTF) (DEXTRANSUCRASE)
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).
 GN GTF.
 OS STREPTOCOCCUS SALIVARIUS.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
 OC STREPTOCOCCUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 25975;
 RX MEDLINE; 93811463.
 RA GIFFARD P.M., ALLEN D.M., MILWARD C.P., SIMPSON C.L., JACQUES N.A.;
 RT "Sequence of the gtf gene of Streptococcus salivarius ATCC 25975 and
 evolution of the gtf genes of oral streptococci."
 RL J. GEN. MICROBIOL. 139:1511-1522(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 25975;
 RX MEDLINE; 92148377.
 RA GIFFARD P.M., SIMPSON C.L., MILWARD C.P., JACQUES N.A.;
 RT "Molecular characterization of a cluster of at least two
 glucosyltransferase genes in Streptococcus salivarius ATCC 25975."
 RL J. GEN. MICROBIOL. 137:2577-2593(1991).
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
 TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE
 BECAUSE OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE
 THE AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
 CC -1- CATALYTIC ACTIVITY: SUCROSE +
 (1,6-ALPHA-D-GLUCOSYL)(N) - D-FRUCTOSE +
 (1,6-ALPHA-D-GLUCOSYL)(N+1).
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -1- DISEASE: DENTAL CARIES.
 CC -1- SIMILARITY: TO REGIONS OF BARLEY AND BACILLUS AMYLOLIQUEFACIENS
 CC ALPHA AMYLASES AND RABBIT GLYCOGEN PHOSPHORYLASE.
 DR EMBL; 211872; G47531; -;
 DR EMBL; 211873; G47528; -;
 DR EMBL; M64111; G153650; -;
 DR PIRAM; PF00128; alpha-amylase; 1.
 KW TRANSFERASE; GLYCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.
 FT SIGNAL 1 42
 FT CHAIN 43 1599
 FT SIMILAR 867 967
 FT SIMILAR 1132 1219
 SQ SEQUENCE 1599 AA; 176480 MW; 575C79C9 CRC32;
 Query Match 68.4%; Score 106; DB 2; Length 1599;
 st Local Similarity 77.3%; Pred. No. 1.09e-09;
 tches 17; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 DB 1402 TGAOVINGHLYFDANGROYKG 1423
 OY 1 TGARTINGOLLYFRANGVOYKG 22

Search completed: Tue Jan 11 15:49:06 2000
 Job time : 52 secs.

 M I S E R I A (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mpsrch_dp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Jan 11 15:37:52 2000; Maspar time 5.64 Seconds

 Output not generated. 203.291 Million cell updates/sec

Title: >US-09-290-049-1

Description: (1-21) from US09290049.pep

Perfect Score: 148

Sequence: 1 ANDHLSILEAWSNDNTPYLHD 21

Scoring table: PAM 150

Searched: 179066 segs, 54579741 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: sptrembl9
 1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
 5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
 9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
 13:sp_vertebrate 14:sp_virus

Statistics: Mean 28.660; Variance 41.580; scale 0.689

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Hit No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	148	100.0	1390	2	069385	GLUCOSYLTRANSFERASE-SI	8.69e-19
2	148	100.0	1455	2	069391	GLUCOSYLTRANSFERASE-SI	8.69e-19
3	148	100.0	1476	2	069390	GLUCOSYLTRANSFERASE-SI	8.69e-19
4	148	100.0	1476	2	069381	GLUCOSYLTRANSFERASE-SI	8.69e-19
5	148	100.0	1476	2	069384	GLUCOSYLTRANSFERASE-SI	8.69e-19
6	148	100.0	1476	2	069396	GLUCOSYLTRANSFERASE-SI	8.69e-19
7	148	100.0	1476	2	069387	GLUCOSYLTRANSFERASE-SI	8.69e-19
8	136	91.9	1455	2	069397	GLUCOSYLTRANSFERASE-SI	7.09e-16
9	136	91.9	1455	2	069388	GLUCOSYLTRANSFERASE-SI	7.09e-16
10	136	91.9	1455	2	069382	GLUCOSYLTRANSFERASE-SI	7.09e-16
11	131	88.5	1590	2	055263	GLUCOSYLTRANSFERASE-SI	1.11e-14
12	131	88.5	1590	2	059983	GLUCOSYLTRANSFERASE-SI	1.11e-14
13	101	68.2	1462	2	069398	GLUCOSYLTRANSFERASE-SI	8.76e-08
14	101	68.2	1462	2	069386	GLUCOSYLTRANSFERASE-SI	8.76e-08
15	101	68.2	1462	2	069389	GLUCOSYLTRANSFERASE-SI	8.76e-08
16	101	68.2	1462	2	069392	GLUCOSYLTRANSFERASE-SI	8.76e-08
17	101	68.2	1462	2	069383	GLUCOSYLTRANSFERASE-SI	8.76e-08
18	101	68.2	1508	2	052224	GLUCOSYLTRANSFERASE-SI	8.76e-08
19	95	64.2	1577	2	054178	GLUCOSYLTRANSFERASE-SI	1.78e-06
20	86	58.1	1449	2	055264	GLUCOSYLTRANSFERASE-SI	1.41e-04

21	86	58.1	1449	2	068542	GLUCOSYLTRANSFERASE N	1.41e-04
22	83	56.1	1577	2	055265	GLUCOSYLTRANSFERASE PR	5.88e-04
23	78	52.7	1599	2	000599	GLUCOSYLTRANSFERASE S	5.88e-04
24	75	50.7	1518	2	000600	GLUCOSYLTRANSFERASE I	2.27e-02
25	72	48.6	579	2	084087	HYPOTHETICAL 65.4 KD P	8.55e-02
26	70	47.3	549	10	043274	T CYTOSOL PLASMA MALE STER	2.03e-01
27	70	47.3	601	2	005566	PPDK.	2.03e-01
28	68	45.9	175	14	P89063	NSP4.	4.77e-01
29	68	45.9	175	14	P89062	NSP4.	4.77e-01
30	67	45.3	490	2	006579	HYPOTHETICAL 52.0 KD P	7.26e-01
31	67	45.3	525	3	059679	HYPOTHETICAL 60.3 KD P	7.26e-01
32	66	44.6	464	10	004073	PHOTOSYSTEM II D1 PROT	1.10e+00
33	66	44.6	4848	2	007944	PRISTINOMYCIN I SYNTHA	1.10e+00
34	65	43.9	581	2	034023	HYPOTHETICAL 66.2 KD P	1.66e+00
35	64	43.2	515	10	023614	PSII D1 PROTEIN PROCES	2.50e+00
36	64	43.2	542	10	P93344	ALDEHYDE DEHYDROGENASE	2.50e+00
37	64	43.2	594	14	091465	CHITINASE.	2.50e+00
38	63	42.6	175	14	056850	NSP4.	3.75e+00
39	63	42.6	175	14	056851	NSP4.	3.75e+00
40	63	42.6	403	10	023222	MADS-BOX PROTEIN HOMOL	3.75e+00
41	62	41.9	360	5	045983	ZC513.10 PROTEIN.	5.59e+00
42	62	41.9	380	14	085633	V-MIL PROTEIN.	5.59e+00
43	62	41.9	420	13	090893	C-MIL PROTO-ONCOGENE (5.59e+00
44	62	41.9	602	11	063485	RAT ACTIVATED C-RAF ON	5.59e+00
45	62	41.9	1009	5	017380	F5263.4 PROTEIN.	5.59e+00

ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	1390 AA.
AC	069385			
AD	069385			
DT	01-AUG-1998 (TREMBLREL. 07, CREATED)			
DT	01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)			
GN	GLUCOSYLTRANSFERASE-SI.			
OS	STREPTOCOCCUS MUTANS.			
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;			
OC	STREPTOCOCCUS.			
NP	SEQUENCE FROM N.A.			
RC	STRAIN-MT4245.			
RC	MEDLINE: 98231643.			
RA	FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,			
RA	KIMURA S., HAMADA S.;			
RT	"Molecular analyses of glucosyltransferase genes among strains of			
RT	Streptococcus mutans."			
RT	FEMS MICROBIOL. LETT. 161:331-336(1998).			
DR	EMBL: D88655; D1027046; -			
DR	TRANSFERASE.			
SO	SEQUENCE 1390 AA; 155375 MW; C211E7B8 CRC32;			

Query Match 100.0%; Score 148; DB 2; Length 1390;
 Best Local Similarity 100.0%; Pred. No. 8.69e-19;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB	507 ANDHLSILEAWSNDNTPYLHD 527
QY	1 ANDHLSILEAWSNDNTPYLHD 21
RESULT	2
ID	069391
AC	069391
DT	01-AUG-1998 (TREMBLREL. 07, CREATED)
DT	01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE	GLUCOSYLTRANSFERASE-SI.
GN	GTRC.
OS	STREPTOCOCCUS MUTANS.
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC	STREPTOCOCCUS.

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MT4251;
 RX MEDLINE: 98231643.
 RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
 RA KIMURA S., HAMADA S.;
 RT "Molecular analyses of glucosyltransferase genes among strains of
 RT Streptococcus mutans.";
 RL FEMS MICROBIOL. LETT. 161:331-336(1998).
 DR EMBL: D88651; D1027054; -
 KW TRANSFERASE.
 SQ SEQUENCE 1455 AA; 162804 MW; DB730514 CRC32;

Query Match 100.0%; Score 148; DB 2; Length 1455;
 Best Local Similarity 100.0%; Pred. No. 8.69e-19;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 507 ANDHLSILEMSDNDTPYLHD 527
 1 ANDHLSILEMSDNDTPYLHD 21

RESULT 3
 ID 069390 PRELIMINARY; PRT: 1476 AA.
 AC 069390;
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE GLUCOSYLTRANSFERASE-I.
 GN GTFB.
 OS STREPTOCOCCUS MUTANS.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
 OC STREPTOCOCCUS.
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MT4251;
 RX MEDLINE: 98231643.
 RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
 RA KIMURA S., HAMADA S.;
 RT "Molecular analyses of glucosyltransferase genes among strains of
 RT Streptococcus mutans.";
 RL FEMS MICROBIOL. LETT. 161:331-336(1998).
 DR EMBL: D88650; D1027053; -
 KW TRANSFERASE.
 SQ SEQUENCE 1476 AA; 165833 MW; 9119672B CRC32;

Query Match 100.0%; Score 148; DB 2; Length 1476;
 Best Local Similarity 100.0%; Pred. No. 8.69e-19;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 481 ANDHLSILEMSDNDTPYLHD 501
 1 ANDHLSILEMSDNDTPYLHD 21

RESULT 4
 ID 069381 PRELIMINARY; PRT: 1476 AA.
 AC 069381;
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE GLUCOSYLTRANSFERASE-I.
 GN GTFB.
 OS STREPTOCOCCUS MUTANS.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
 OC STREPTOCOCCUS.
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MT8148;
 RX MEDLINE: 98231643.
 RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
 RA KIMURA S., HAMADA S.;
 RT "Molecular analyses of glucosyltransferase genes among strains of

RT Streptococcus mutans.";
 RL FEMS MICROBIOL. LETT. 161:331-336(1998).
 DR EMBL: D88651; D1027041; -
 KW TRANSFERASE.
 SQ SEQUENCE 1476 AA; 165638 MW; 0868C20A CRC32;

Query Match 100.0%; Score 148; DB 2; Length 1476;
 Best Local Similarity 100.0%; Pred. No. 8.69e-19;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 481 ANDHLSILEMSDNDTPYLHD 501
 1 ANDHLSILEMSDNDTPYLHD 21

RESULT 5
 ID 069384 PRELIMINARY; PRT: 1476 AA.
 AC 069384;
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE GLUCOSYLTRANSFERASE-I.
 GN GTFB.
 OS STREPTOCOCCUS MUTANS.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
 OC STREPTOCOCCUS.
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MT4245;
 RX MEDLINE: 98231643.
 RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
 RA KIMURA S., HAMADA S.;
 RT "Molecular analyses of glucosyltransferase genes among strains of
 RT Streptococcus mutans.";
 RL FEMS MICROBIOL. LETT. 161:331-336(1998).
 DR EMBL: D88654; D1027045; -
 KW TRANSFERASE.
 SQ SEQUENCE 1476 AA; 165792 MW; C76D8483 CRC32;

Query Match 100.0%; Score 148; DB 2; Length 1476;
 Best Local Similarity 100.0%; Pred. No. 8.69e-19;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 481 ANDHLSILEMSDNDTPYLHD 501
 1 ANDHLSILEMSDNDTPYLHD 21

RESULT 6
 ID 069396 PRELIMINARY; PRT: 1476 AA.
 AC 069396;
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE GLUCOSYLTRANSFERASE-I.
 GN GTFB.
 OS STREPTOCOCCUS MUTANS.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
 OC STREPTOCOCCUS.
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MT4467;
 RX MEDLINE: 98231643.
 RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
 RA KIMURA S., HAMADA S.;
 RT "Molecular analyses of glucosyltransferase genes among strains of
 RT Streptococcus mutans.";
 RL FEMS MICROBIOL. LETT. 161:331-336(1998).
 DR EMBL: D89977; D1027059; -
 KW TRANSFERASE.
 SQ SEQUENCE 1476 AA; 165647 MW; 258D4ADD CRC32;

Query Match 100.0%; Score 148; DB 2; Length 1476;

Best Local Similarity 100.0%; Pred. No. 8.69e-19;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 481 ANDHLSILEAMSDNDTPYLHD 501
OY 1 ANDHLSILEAMSDNDTPYLHD 21

RESULT 7
ID 069387 PRELIMINARY; PRT; 1476 AA.
AC 069387;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE GLUCOSYLTRANSFERASE-I.
GN GTFB.
OS STREPTOCOCCUS MUTANS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MT4239;
RX MEDLINE: 98231643.
RA FUJIWARA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
RA KIMURA S., HAMADA S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT streptococcus mutans."
RL FEMS MICROBIOL. LETT. 161:331-336(1998).
DR EMBL: D88657; D1027049; -.
KW TRANSFERASE.
SQ SEQUENCE 1476 AA; 165819 MW; AF6DBA6F CRC32;

Query Match
Best Local Similarity 100.0%; Score 148; DB 2; Length 1476;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 481 ANDHLSILEAMSDNDTPYLHD 501
OY 1 ANDHLSILEAMSDNDTPYLHD 21

RESULT 8
ID 069397 PRELIMINARY; PRT; 1455 AA.
AC 069397;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE GLUCOSYLTRANSFERASE-SI.
GN GTFB.
OS STREPTOCOCCUS MUTANS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MT4467;
RX MEDLINE: 98231643.
RA FUJIWARA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
RA KIMURA S., HAMADA S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT streptococcus mutans."
RL FEMS MICROBIOL. LETT. 161:331-336(1998).
DR EMBL: D89978; D1027060; -.
KW TRANSFERASE.
SQ SEQUENCE 1455 AA; 162914 MW; 7D729DA0 CRC32;

Query Match
Best Local Similarity 95.2%; Score 136; DB 2; Length 1455;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 507 ANDHLSILEAMSDNDTPYLHD 527
OY 1 ANDHLSILEAMSDNDTPYLHD 21

RESULT 9
ID 069388 PRELIMINARY; PRT; 1455 AA.
AC 069388;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE GLUCOSYLTRANSFERASE-SI.
GN GTFB.
OS STREPTOCOCCUS MUTANS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MT4239;
RX MEDLINE: 98231643.
RA FUJIWARA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
RA KIMURA S., HAMADA S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT streptococcus mutans."
RL FEMS MICROBIOL. LETT. 161:331-336(1998).
DR EMBL: D88658; D1027050; -.
KW TRANSFERASE.
SQ SEQUENCE 1455 AA; 163046 MW; 7659CFBC CRC32;

Query Match
Best Local Similarity 91.9%; Score 136; DB 2; Length 1455;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 507 ANDHLSILEAMSDNDTPYLHD 527
OY 1 ANDHLSILEAMSDNDTPYLHD 21

RESULT 10
ID 069382 PRELIMINARY; PRT; 1455 AA.
AC 069382;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE GLUCOSYLTRANSFERASE-SI.
GN GTFB.
OS STREPTOCOCCUS MUTANS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MT8148;
RX MEDLINE: 98231643.
RA FUJIWARA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOBUE S.,
RA KIMURA S., HAMADA S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT streptococcus mutans."
RL FEMS MICROBIOL. LETT. 161:331-336(1998).
DR EMBL: D88652; D1027042; -.
KW TRANSFERASE.
SQ SEQUENCE 1455 AA; 162970 MW; 40B022BD CRC32;

Query Match
Best Local Similarity 95.2%; Score 136; DB 2; Length 1455;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 507 ANDHLSILEAMSDNDTPYLHD 527
OY 1 ANDHLSILEAMSDNDTPYLHD 21

RESULT 11
ID 055263 PRELIMINARY; PRT; 1590 AA.
AC 055263;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)

DE GLUCOSYLTRANSFERASE GTF-I.
 OS STREPTOCOCCUS SOBRINUS.
 OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
 OC STREPTOCOCCUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 33478;
 RA SATO S.;
 RL ANN. KAGOSHIMA UNIV. DENTAL SCHOOL 16:23-29(1996).
 DR EMBL; D63570; D1010438;
 PFAM; PF00128; alpha-amylase; 1.
 KW TRANSFERASE.
 SQ SEQUENCE 1590 AA; 176057 MW; 47186F6A CRC32;
 Query Match 88.5%; Score 131; DB 2; Length 1590;
 Best Local Similarity 85.7%; Pred. No. 1,11e-14;
 Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

477 ANNHVSIVEAWSNDTPYLHD 497
 1 ANDHLSILEAWSNDTPYLHD 21

RESULT 12
 ID 059983; PRELIMINARY; PRT; 1590 AA.
 AC 059983;
 DT 01-NOV-1996 (TREMBLEREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLEREL. 01, LAST SEQUENCE UPDATE)
 DT 01-JAN-1999 (TREMBLEREL. 09, LAST ANNOTATION UPDATE)
 DE GLUCOSYLTRANSFERASE-1 PRECURSOR (EC 2.4.1.5) (DETRANSGRASE) (SUCROSE
 DE 6-GLUCOSYLTRANSFERASE).
 GN GTF.
 OS STREPTOCOCCUS SOBRINUS.
 OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
 OC STREPTOCOCCUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OMZ176;
 RX MEDLINE; 94146405.
 RA SATO S., INOUE M., HANDA N., AIZAWA Y., ISOBE Y., KATAYAMA T.;
 RT "DNA sequence of the glucosyltransferase gene of serotype d
 RT Streptococcus sobrinus".
 RL DNA SEQ. 4:19-27(1993).
 CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -
 CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).
 CC EMBL; D13858; D1003482;
 DR PFAM; PF00128; alpha-amylase; 1.
 H SIGNAL; TRANSFERASE; GLUCOSYLTRANSFERASE.
 FT SIGNAL 38 POTENTIAL.
 FT CHAIN 39 1590 GLUCOSYLTRANSFERASE-1.
 SQ SEQUENCE 1590 AA; 175955 MW; F65A57D7 CRC32;
 Query Match 88.5%; Score 131; DB 2; Length 1590;
 Best Local Similarity 85.7%; Pred. No. 1,11e-14;
 Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 477 ANNHVSIVEAWSNDTPYLHD 497
 1 ANDHLSILEAWSNDTPYLHD 21

RESULT 13
 ID 069398; PRELIMINARY; PRT; 1462 AA.
 AC 069398;
 DT 01-AUG-1998 (TREMBLEREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLEREL. 07, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLEREL. 08, LAST ANNOTATION UPDATE)
 DE GLUCOSYLTRANSFERASE-S.
 GN GTFD.
 OS STREPTOCOCCUS MUTANS.
 OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
 OC STREPTOCOCCUS.
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-MT4467;
 RX MEDLINE; 98231643.
 RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOUBE S.,
 RA KIMURA S., HAMADA S.;
 RT "Molecular analyses of glucosyltransferase genes among strains of
 RT Streptococcus mutans".
 RL FEMS MICROBIOL. LETT. 161:331-336(1998).
 DR EMBL; D89979; D1027061;
 KW TRANSFERASE.
 SQ SEQUENCE 1462 AA; 163267 MW; E9BD264F CRC32;
 Query Match 68.2%; Score 101; DB 2; Length 1462;
 Best Local Similarity 77.8%; Pred. No. 8.76e-08;
 Matches 14; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 495 AINHLSILEAWSNDTPY 512
 1 ANDHLSILEAWSNDTPY 18

RESULT 14
 ID 069386; PRELIMINARY; PRT; 1462 AA.
 AC 069386;
 DT 01-AUG-1998 (TREMBLEREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLEREL. 07, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLEREL. 08, LAST ANNOTATION UPDATE)
 DE GLUCOSYLTRANSFERASE-S.
 GN GTFD.
 OS STREPTOCOCCUS MUTANS.
 OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
 OC STREPTOCOCCUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MT4245;
 RX MEDLINE; 98231643.
 RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOUBE S.,
 RA KIMURA S., HAMADA S.;
 RT "Molecular analyses of glucosyltransferase genes among strains of
 RT Streptococcus mutans".
 RL FEMS MICROBIOL. LETT. 161:331-336(1998).
 DR EMBL; D88656; D1027047;
 KW TRANSFERASE.
 SQ SEQUENCE 1462 AA; 163464 MW; F37E5F30 CRC32;
 Query Match 68.2%; Score 101; DB 2; Length 1462;
 Best Local Similarity 77.8%; Pred. No. 8.76e-08;
 Matches 14; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 495 AINHLSILEAWSNDTPY 512
 1 ANDHLSILEAWSNDTPY 18

RESULT 15
 ID 069389; PRELIMINARY; PRT; 1462 AA.
 AC 069389;
 DT 01-AUG-1998 (TREMBLEREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLEREL. 07, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLEREL. 08, LAST ANNOTATION UPDATE)
 DE GLUCOSYLTRANSFERASE-S.
 GN GTFD.
 OS STREPTOCOCCUS MUTANS.
 OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
 OC STREPTOCOCCUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MT4239;
 RX MEDLINE; 98231643.
 RA FUJIMURA T., TERAO Y., HOSHINO T., KAWABATA S., OOSHIMA T., SOUBE S.,
 RA KIMURA S., HAMADA S.;
 RT "Molecular analyses of glucosyltransferase genes among strains of
 RT Streptococcus mutans".

Thu Jan 13 09:13:35 2000

US-09-290-049-1.rsp

Page 5

RL FEMS MICROBIOL. LETT. 161:331-336(1998).
DR EMBL: D88659; D1027051; -.
SQ SEQUENCE 1462 AA; 163554 MW; B057EF28 CRC32;

Query Match	68.2%;	Score 101;	DB 2;	Length 1462;
Best Local Similarity	77.8%;	Pred. No. 8.76e-08;		
Matches	14;	Conservative	1;	Mismatches 3; Indels 0; Gaps 0;

```

Db      495 AINHL$ILEAWSNDNDPQY 512
      | :| | | | | | | | | |
QY      1 ANDHLS$ILEAWSNDNDTPY 18

```

Search completed: Tue Jan 11 15:38:39 2000
Job time : 47 secs.

 W I N D O W
 ***** (TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Msrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Tue Jan 11 15:37:26 2000; Maspar time 2.42 Seconds
 245.606 Million cell updates/sec
 No output not generated.

Title: >US-09-290-049-1
 Description: (1-21) from US09290049.pep
 Perfect Score: 148
 Sequence: 1 ANDHLSTLEAMSDNDPYLHD 21

Scoring table: PAM 150
 Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: swiss-prot37
 1:swissprot

Statistics: Mean 29.534; Variance 43.172; scale 0.684

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description	Pred. No.
1	148	100.0	1475	1 GTFB_STRMU GLUCOSYLTRANSFERASE-I	1.49e-18
2	136	91.9	1375	1 GTFB_STRMU GLUCOSYLTRANSFERASE-SI	1.10e-15
3	131	88.5	1592	1 GTF2_STRDO GLUCOSYLTRANSFERASE-I	1.66e-14
4	131	88.5	1597	1 GTF1_STRDO GLUCOSYLTRANSFERASE-I	1.66e-14
5	102	68.2	1365	1 GTF5_STRDO GLUCOSYLTRANSFERASE-S	9.98e-08
6	101	68.2	1430	1 GTFD_STRMU GLUCOSYLTRANSFERASE-S	7.25e-02
7	72	48.6	330	1 G3PC_LEIME GLYCERALDEHYDE 3-PHOSP	7.25e-02
8	72	48.6	336	1 UL16_EBY BGLF2 PROTEIN.	7.25e-02
9	66	44.6	1885	1 FAS2_CANAL FATTY ACID SYNTHASE, S	8.72e-01
10	64	43.2	476	1 PCHA_PSEAE SALICYLATE BIOSYNTHET	1.94e+00
11	64	43.2	480	1 DISA_TRIGA PUTATIVE VENOM METALLO	1.94e+00
12	64	43.2	638	1 KRAF_XENLA RAF PROTO-ONCOGENE SER	1.94e+00
13	63	42.6	1131	1 PMAI_DUNBI PLASMA MEMBRANE ATPASE	4.23e+00
14	62	41.9	323	1 KRAF_MSV36 RAF SERINE/THREONINE-P	4.23e+00
15	62	41.9	380	1 KRAF_AVIMH MIL SERINE/THREONINE-P	4.23e+00
16	62	41.9	546	1 SYL_PYRFU ISOLEUCYL-TRNA SYNTHET	4.23e+00
17	62	41.9	647	1 KMIT_CHICK MIL PROTO-ONCOGENE SER	4.23e+00
18	62	41.9	648	1 KRAF_RAT RAF PROTO-ONCOGENE SER	4.23e+00
19	62	41.9	648	1 KRAF_HUMAN RAF PROTO-ONCOGENE SER	4.23e+00
20	61	41.2	181	1 SYDE_ECOLI SYD PROTEIN.	6.21e+00
21	61	41.2	202	1 HR2A_TRIFL HEMORRHAGIC METALLOPRO	6.21e+00
22	61	41.2	330	1 G3PC_TRYBB GLYCERALDEHYDE 3-PHOSP	6.21e+00
23	61	41.2	330	1 G3PI_ECOLI GLYCERALDEHYDE 3-PHOSP	6.21e+00

ALIGNMENTS

RESULT ID	1	STANDARD:	PRT:	1475 AA.
AC	P08987:			
AD	01-NOV-1988 (REL. 09, CREATED)			
DT	01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)			
DT	15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)			
DE	GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-I) (DEXTRANSUCRASE)			
GN	GTFB.			
OS	STREPTOCOCCUS MUTANS.			
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;			
OC	STREPTOCOCCUS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-GS-5:			
RX	MEDLINE: 87308013.			
RA	SHIROZA T., UEDA S., KURAMITSU H.K.;			
RT	"Sequence analysis of the gtfb gene from Streptococcus mutans.";			
RL	J. BACTERIOL. 169:4263-4270(1987).			
CC	- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.			
CC	- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).			
CC	- D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).			
CC	- SUBCELLULAR LOCATION: SECRETED.			
CC	- DISEASE: DENTAL CARIES.			
CC	- GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF GLUCANS.			
CC	- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-BINDING PROTEIN FROM S. MUTANS.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by, and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL, M17361; G153640; -			
DR	PIR: B33135; B33135.			
DR	PFAM: PF00128; alpha-amylase; 1.			
KW	TRANSFERASE; GLUCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.			
FT	SIGNAL 1 34 POTENTIAL.			

```

FT CHAIN 35 1475 GLUCOSYLTRANSFERASE-1.
FT DOMAIN 35 1050 CATALYTIC (APPROXIMATE).
FT TRANSFERASE 1096 1475 GLUCAN-BINDING (APPROXIMATE).
FT REPEAT 1096 1129 A REPEAT.
FT DOMAIN 1160 1469 5 X TANDEM REPEATS.
FT REPEAT 1160 1209 1.
FT REPEAT 1224 1274 2.
FT REPEAT 1289 1339 3.
FT REPEAT 1354 1404 4.
FT REPEAT 1419 1469 5.
FT REPEAT 1475 165812 MM; 4542C1D6 CRC32;
SQ SEQUENCE 1475 AA; 165812 MM; 4542C1D6 CRC32;

Query Match 100.0%; Score 148; DB 1; Length 1475;
Best Local Similarity 100.0%; Pred. No. 1,49e-18;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 481 ANDHLSILEMSDNDPTPLD 501
1 ANDHLSILEMSDNDPTPLD 21

RESULT 2
ID GTF2_STRMU STANDARD; PRT; 1375 AA.
AC P13470; P05427;
DT 01-NOV-1988 (REL. 09, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GLUCOSYLTRANSFERASE-SI PRECURSOR (EC 2.4.1.5) (GTF-SI)
DE (DEXTRANSUCRASE) (SUCROSE 6-GLUCOSYLTRANSFERASE).
GN GTF.
OS STREPTOCOCCUS MUTANS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-GS-5;
RX MEDLINE: 89137980.
RX UEDA S., SHIROZA T., KURAMITSU H.K.;
RT "Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.";
RL GENE 69:101-109(1988).
RN [2]
RP SEQUENCE OF 1-349 FROM N.A.
RX STRAIN-GS-5;
RX MEDLINE: 87308013.
RX SHIROZA T., UEDA S., KURAMITSU H.K.;
RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";
RL J. BACTERIO. 169:4263-4270(1987).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) -
CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DISEASE: DENTAL CARIES.
CC -1- GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED
CC GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE
CC GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF
CC GLUCANS.
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S. MUTANS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: M22054; G153643; -
CC DR EMBL: M17361; G153641; -
CC PIR: J70345; J70345.

```

```

DR PIR: C33135; C33135.
DR PFAM: PF00128; alpha-amylase; 1.
DR TRANSFERASE 1096 1475 GLUCAN-BINDING (APPROXIMATE).
DR REPEAT 1096 1129 A REPEAT.
DR DOMAIN 1160 1469 5 X TANDEM REPEATS.
DR REPEAT 1160 1209 1.
DR REPEAT 1224 1274 2.
DR REPEAT 1289 1339 3.
DR REPEAT 1354 1404 4.
DR REPEAT 1419 1469 5.
DR REPEAT 1475 165812 MM; 4542C1D6 CRC32;
SQ SEQUENCE 1475 AA; 165812 MM; 4542C1D6 CRC32;

Query Match 91.9%; Score 136; DB 1; Length 1375;
Best Local Similarity 95.2%; Pred. No. 1.10e-15;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 507 ANDHLSILEMSDNDPTPLD 527
1 ANDHLSILEMSDNDPTPLD 21

RESULT 3
ID GTF2_STRDO STANDARD; PRT; 1592 AA.
AC P27470;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GLUCOSYLTRANSFERASE-1 PRECURSOR (EC 2.4.1.5) (GTF-1) (DEXTRANSUCRASE)
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).
OS STREPTOCOCCUS DORNI (STREPTOCOCCUS SOBRINUS).
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-6715;
RX MEDLINE: 91123227.
RX ABO H., MATSUMURA T., KODAMA T., OHTA H., FUKUI K., KATO K.,
RT "Peptide sequences for sucrose splitting and glucan binding within
RT Streptococcus sobrinus glucosyltransferase (water-insoluble glucan
RT synthetase).";
RL J. BACTERIO. 173:989-996(1991).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) -
CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DISEASE: DENTAL CARIES.
CC -1- GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED
CC GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE
CC GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF
CC GLUCANS.
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
CC BINDING PROTEIN FROM S. MUTANS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: D90213; G217033; -
CC DR EMBL: A38175; A38175.
CC DR PFAM: PF00128; alpha-amylase; 1.
CC HSSP: P00695; ZHEB.
CC TRANSFERASE: GLUCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.

```


FT REPEAT 1225 1274 3.
FT REPEAT 1289 1339 4.
FT REPEAT 1353 1365 5 (INCOMPLETE).
SQ SEQUENCE 1365 AA; 151590 MW; 156F8B31 CRC32;

Query Match 68.9%; Score 102; DB 1; Length 1365;
Best Local Similarity 71.4%; Pred. No. 6,05e-08;
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

DB 467 AIDHLSILEAMSGNDNDYK 487
1 AIDHLSILEAMSGNDNDYK 21

RESULT 6
ID GTFD_STRMU STANDARD; PRT; 1430 AA.
AC P49331;
DT 01-FEB-1996 (REL. 33, CREATED)
01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
01-NOV-1997 (REL. 33, LAST SEQUENCE UPDATE)
DE (SUCCOSYLTRANSFERASE-S PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)
GN GTFD.
OS STREPTOCOCCUS MUTANS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-GS-5;
RX MEDLINE: 91100958.
RA HONDA O., KATO C., KURAMITSU H.K.;
RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding
the glucosyltransferase-S enzyme."
RL J. GEN. MICROBIOL. 136:2099-2105(1990).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) =
D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DISEASE: DENTAL CARIES.
CC -1- GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED
GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE
GLUCANS (ALPHA 1,6-GLUCOSE). GTF-S1 SYNTHESIZES BOTH FORMS OF
GLUCANS.
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
BINDING PROTEIN FROM S. MUTANS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL: M29296; G153645; -
DR PFAM: PF00128; alpha-amylase; 1.
KM TRANSFERASE; GLUCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.
FT SIGNAL 1 1430
FT CHAIN ? 1430
FT DOMAIN 1232 1423 3 X 63 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 1232 1295 1.
FT REPEAT 1296 1359 2.
FT REPEAT 1360 1423 3.
SQ SEQUENCE 1430 AA; 159765 MW; D3DE3681 CRC32;

QY 1 AIDHLSILEAMSGNDNDYK 18

RESULT 7
ID G3PC_LETME STANDARD; PRT; 330 AA.
AC 001558;
DT 01-APR-1993 (REL. 25, CREATED)
01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE, CYTOSOLIC (EC 1.2.1.12)
GN GAPDH.
OS LEISHMANIA MEXICANA.
OC EUKARYOTA; EUKLENOZOA; KINETOPLASTIDA; TRYPAANOMATIDAE; LEISHMANIA.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-SSP. MEXICANA;
RX MEDLINE: 93063042.
RA HANNAERT V., BLAAUW M., KOHL L., ALLERT S., OPPERDOES F.R.,
RA MICHELIS P.A.M.;
RT "Molecular analysis of the cytosolic and glycosomal glyceraldehyde-3-
phosphate dehydrogenase in Leishmania mexicana."
RL MOL. BIOCHEM. PARASITOL. 55:115-126(1992).
CC -1- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE
+ NAD(+) = 1,3-DIPHOSPHATEGLYCERATE + NADH.
CC -1- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL: X65220; G9553; -
DR PIR: S25142; S25142.
DR PIR: B48445; B48445.
DR PROSITE: PS00071; GAPDH; 1.
DR PFAM: PF00044; gpdh; 1.
DR HSSP: P06977; IGAD.
KM GLYCOLYSIS; OXIDOREDUCTASE; NAD.
FT INIT MET 0
FT BINDING 148 148
FT ACT_SITE 175 175
SQ SEQUENCE 330 AA; 35511 MW; 80B6A7E9 CRC32;

Query Match 48.6%; Score 72; DB 1; Length 330;
Best Local Similarity 52.6%; Pred. No. 7,25e-02;
Matches 10; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

DB 300 NDHFVKLVSYNDYNGTSH 318
2 NDHLSILEAMSGNDNDYK 20

QY 2 NDHLSILEAMSGNDNDYK 20

RESULT 8
ID U116_EBV STANDARD; PRT; 336 AA.
AC P03221;
DT 21-JUL-1986 (REL. 01, CREATED)
21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DE BGLF2 PROTEIN.
GN BGLF2.
OS EPSTEIN-BARR VIRUS (STRAIN B95-8) (HUMAN HERPESVIRUS 4).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC GAMMAHERPESVIRINAE; LYMPHOCRYPTOVIRUS.
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE: 84270667.

RA BAER R., BANKER A.T., BIGGIN M.D., DEININGER P.L., FARRELL P.J.,
 RA GIBSON T.J., HATFIELD G., HUDSON G.S., SATCHWELL S.C., SEGGIN C.,
 RA TUFENELL P.S., BARRELL B.G.,
 RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.",
 RL NATURE 310:207-211(1984).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 92113548.
 RA CHEN M.R., HSU T.Y., LIN S.W., CHEN J.Y., YANG C.S.,
 RT "Cloning and characterization of cDNA clones corresponding to
 RT transcripts from the BamHI G region of the Epstein-Barr virus genome
 RT and expression of BGLF2.",
 RL J. GEN. VIROL. 72:3047-3055(1991).
 CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL16,
 CC HSV-6 ORF1R, EBV-1 46, HCMV UL94, EBV BDLF2, AND VZV 44.
 CC -
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M60514; G330339; -
 CC EMBL: V01555; G59136; -
 CC EMBL: S77132; G243315; -
 CC PIR: A03784; Q0BE40.
 CC PIR: S33036; S33036.
 CC LATE PROTEIN.
 CC SEQUENCE 336 AA; 36888 MW; 2DDDD25B CRC32;
 SO
 Query Match 48.6%; Score 72; DB 1; Length 336;
 Best Local Similarity 41.2%; Pred. No. 7,25e-02;
 Matches 7; Conservative 7; Mismatches 3; Indels 0; Gaps 0;
 DB 216 AGARVNLRCGTEDDSP 232
 QY 1 ANDHLSTLEAMSDNDTP 17
 ID FAS2-CANAL STANDARD; PRT; 1885 AA.
 AC P43098;
 DT 01-NOV-1995 (REL. 32, CREATED)
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE FATTY ACID SYNTHASE, SUBUNIT ALPHA (EC 2.3.1.86) [CONTAINS:
 DE EC 1.1.1.100; EC 2.3.1.41].
 DE FAS2.
 DE CANDIDA ALBICANS (YEAST).
 CC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCAROMYCETALES;
 CC CANDIDACEAE; CANDIDA.
 CC [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-4918;
 RX MEDLINE: 95255657.
 RA SOUTHWARD S.B., CIHAR R.L.,
 RT "Analysis and expression of the Candida albicans FAS2 gene.",
 RL GENE 156:133-138(1995).
 CC -1- FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
 CC LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
 CC THE ALPHA SUBUNIT CONTAINS DOMAINS FOR: ACYL-CARRIER PROTEIN,
 CC 3-OXOACYL-(ACYL-CARRIER PROTEIN) REDUCTASE, AND 3-OXOACYL-(ACYL-
 CC CARRIER-PROTEIN) SYNTHASE.
 CC -1- CATALYTIC ACTIVITY: ACETYL-COA + N MALONYL-COA + 2N NADPH =
 CC LONG-CHAIN FATTY ACID + (N+1) COA + N CO(2) + 2N NADP(+).
 CC -1- CATALYTIC ACTIVITY: ACYL-[ACYL-CARRIER PROTEIN] + MALONYL-[ACYL-
 CC CARRIER PROTEIN] = 3-OXOACYL-[ACYL-CARRIER PROTEIN] + CO(2) +
 CC [ACYL-CARRIER PROTEIN].
 CC -1- CATALYTIC ACTIVITY: (3R)-3-HYDROXYACYL-[ACYL-CARRIER PROTEIN] +
 CC NADP(+) = 3-OXOACYL-[ACYL-CARRIER PROTEIN] + NADPH.
 CC -1- SUBUNIT: FATTY ACID SYNTHETASE ARE [ALPHA(6)BETA(6)] HEXAMERS OF

CC TWO MULTIFUNCTIONAL SUBUNITS (ALPHA & BETA).
 CC -1- SIMILARITY: TO THE FATTY ACID SYNTHETASE, SUBUNIT ALPHA FROM
 CC OTHER FUNGI.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L29063; G456443; -
 CC DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 1.
 CC DR PROSITE: PS00606; B-KETOACYL SYNTHASE; 1.
 CC KM FATTY ACID BIOSYNTHESIS: MULTIFUNCTIONAL ENZYME; OXIDOREDUCTASE;
 CC TRANSFERASE; NADP; PHOSPHOPANTETHEINE.
 CC FT DOMAIN 1 ?
 CC FT DOMAIN 2 ?
 CC FT DOMAIN 3 ?
 CC FT BINDING 181 181 BETA-KETOACYL SYNTHASE.
 CC FT ACT SITE 1304 1304 PHOSPHOPANTETHEINE (BY SIMILARITY).
 CC FT ACT SITE 1304 1304 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
 CC SEQUENCE 1885 AA; 207588 MW; A19558C4 CRC32;
 SO
 Query Match 44.6%; Score 66; DB 1; Length 1885;
 Best Local Similarity 25.0%; Pred. No. 8,72e-01;
 Matches 5; Conservative 8; Mismatches 7; Indels 0; Gaps 0;
 DB 1452 NKRLPQKSWETELSLQE 1471
 QY 2 NDHLSTLEAMSDNDTPYLD 21
 ID PCNA-PSEAE STANDARD; PRT; 476 AA.
 AC Q51508;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE SALICYLATE BIOSYNTHESIS ISOCHORISMATE SYNTHASE (EC 5.4.99.6).
 DE PCNA.
 DE PSEUDOMONAS AERUGINOSA.
 CC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;
 CC PSEUDOMONAS.
 CC [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-ATCC 15692 / PA01;
 RX MEDLINE: 96086939.
 RA SERINO L., REIMANN C., BAUR H., BEYELER M., VISCA P., HAAS D.,
 RT "Structural genes for salicylate biosynthesis from chorismate in
 RT Pseudomonas aeruginosa.",
 RL MOL. GEN. GENET. 249:217-228(1995).
 CC -1- FUNCTION: INVOLVED IN THE CONVERSION OF CHORISMATE TO SALICYLATE
 CC (PROBABLE).
 CC -1- CATALYTIC ACTIVITY: CHORISMATE = ISOCHORISMATE.
 CC -1- PATHWAY: SALICYLATE BIOSYNTHESIS.
 CC -1- SIMILARITY: STRONG, TO OTHER ISOCHORISMATE SYNTHASES; WEAK, TO
 CC TRPE AND PABB.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X82644; G929780; -
 CC DR PRAM: PR00425; chorismate_bind; 1.
 CC KM ISOMERASE.
 CC SEQUENCE 476 AA; 52071 MW; DEDC2D5C CRC32;
 SO
 Query Match 43.2%; Score 64; DB 1; Length 476;

Best Local Similarity 47.18; Pred. No. 1.94e+00;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Db 44 DPLVFGAMDDROTPTCL 60
1 1 1 1 1 1 1 1 1 1
QY 3 DPLVFGAMDDROTPTCL 19

RESULT 11
AC DISA-TRIGA STANDARD; PRT: 480 AA.
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE PUTATIVE VENOM METALLOPROTEINASE PRECURSOR (EC 3.4.24.-) [CONTAINS:
DE DISINTEGRIN TRIGRAMIN ALPHA (PLATELET AGGREGATION ACTIVATION
DE INHIBITOR)]
OS TRIMERESURUS GRAMINEUS (INDIAN GREEN TREE VIBER) (GREEN HADU SNAKE).
EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; LEPIDOSAURIA; SQUAMATA;
SCLEROGLOSSA; SERPENTES; COLUBROIDEA; VIPERIDAE; CROTALINAE;
TRIMERESURUS.
[1]
RN SEQUENCE FROM N.A.
RC TISSUE-VENOM GLAND.
RX MEDLINE: 90332429.
RA NEPPER M.P., JACOBSON M.A.;
*Sequence of a cDNA encoding the platelet aggregation inhibitor
trigramin.*
RT NUCLEIC ACIDS RES. 18:4255-4255(1990).
RN
[2]
RN SEQUENCE OF 408-479.
RX MEDLINE: 89229063.
RA HUANG T.-F., HOLT J.C., KIRBY E.P., NIEMIROWSKI S.;
*Trigramin: primary structure and its inhibition of von Willebrand
factor binding to glycoprotein IIb/IIIa complex on human platelets.*;
RT BIOCHEMISTR 28:661-666(1989).
RN
[3]
RN SEQUENCE OF 408-479.
RC TISSUE-VENOM;
RX MEDLINE: 90207217.
RA DENNIS M.S., HENZEL W.J., PITTI R.M., LIPARI M.T., NAPIER M.A.,
DEISHER T.A., BUNTING S., LAZARUS R.A.;
*Platelet glycoprotein IIb-IIIa protein antagonists from snake
venoms: evidence for a family of platelet-aggregation inhibitors.*;
RT PROC. NATL. ACADE. SCI. U.S.A. 87:2471-2475(1990).
RN
[4]
RN SEQUENCE OF 408-419.
RX MEDLINE: 88058981.
RA HUANG T.-F., HOLT J.C., LUKASIEWICZ H., NIEMIROWSKI S.;
*Trigramin. A low molecular weight peptide inhibiting fibrinogen
interaction with platelet receptors expressed on glycoprotein
IIb-IIIa complex.*;
RT J. BIOL. CHEM. 262:16157-16163(1987).
RN
-1- FUNCTION: THIS PROTEIN IS A ZINC PROTEASE FROM SNAKE VENOM THAT
ACTS IN HEMORRHAGE (BY SIMILARITY).
-1- FUNCTION: TRIGRAMIN INHIBITS FIBRINOGEN INTERACTION WITH PLATELET
RECEPTORS EXPRESSED ON GLYCOPROTEIN IIb-IIIa COMPLEX. ACT BY
BINDING TO THE GLYCOPROTEIN IIb-IIIa RECEPTOR ON THE PLATELET
SURFACE AND INHIBITS AGGREGATION INDUCED BY ADP, THROMBIN,
PLATELET-ACTIVATING FACTOR AND COLLAGEN.
-1- COFACTOR: BINDS AND REQUIRES A ZINC ATOM, WHICH IS ESSENTIAL FOR
PROTEOLYTIC ACTIVITY.
-1- SIMILARITY: BELONGS TO THE REPROLYSIN SUBFAMILY.
-1- SIMILARITY: HIGH, IN THE C-TERMINAL REGION, TO OTHER DISINTEGRINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: X51530; G64408; -
DR PIR: A29784; A29784.
DR PIR: A30065; A30065.
DR PROSITE: PS00142; ZINC_PROTEASE. 1.
DR PROSITE: PS00427; DISINTEGRINS; 1.
DR PFAM: PF00099; zn-protease; 1.
DR PFAM: PF00200; disintegrin; 1.
DR HSP: P18619; LFV.
KW HYDROLASE; METALLOPROTEINASE; ZINC; BLOOD COAGULATION; PLATELET;
KW CELL ADHESION; VENOM; ZYMOGEN; SIGNAL.
FT SIGNAL 1
FT PROPEP 190
FT CHAIN 191 407
FT CHAIN 408 480
FT METAL 333 333
FT ACT_SITE 334 334
FT METAL 337 337
FT METAL 343 343
FT METAL 348 348
FT DISULFID 348 353
FT DISULFID 413 428
FT DISULFID 415 423
FT DISULFID 422 445
FT DISULFID 436 442
FT DISULFID 441 466
FT DISULFID 454 473
FT SITE 458 460
FT CARBOHYD 279 279
SQ SEQUENCE 480 AA; 53494 MW; D59DA91F CRC32;
Query Match 43.28; Score 64; DB 1; Length 480;
Best Local Similarity 63.68; Pred. No. 1.94e+00;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 244 LSVLEIEMSDND 254
QY 5 LSVLEIEMSDND 15
RESULT 12
ID KNAE_XENIA STANDARD; PRT: 638 AA.
AC P09560; Q91390.
DT 01-MAR-1989 (REL. 10, CREATED)
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE RAF PROTO-ONCOGENE SERINE/THREONINE-PROTEIN KINASE (EC 2.7.1.-).
GN C-RAF.
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC MESOBATRACHIA; PIPOIDEA; PIPOIDAE; XENOPODINAE; XENOPUS.
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE-OOCYTE;
RX MEDLINE: 89057471.
RA LE GUELLEC R., LE GUELLEC K., PARIS J., PHILIPPE M.;
Nucleotide sequence of Xopus C-raf coding region.;
RT NUCLEIC ACIDS RES. 16:10357-10357(1988).
RN
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE: 92096753.
RA LE GUELLEC R., COOTURLER A., LE GUELLEC K., PARIS J., LE FUR N.,
RA PHILIPPE M.;
*Xopus C-raf proto-oncogene: cloning and expression during
oogenesis and early development.*;
RT BIOL. CELL 72:39-45(1991).
RN
-1- FUNCTION: INVOLVED IN THE TRANSDUCTION OF MITOGENIC SIGNALS FROM
CC THE CELL MEMBRANE TO THE NUCLEUS. PART OF THE RAS-DEPENDENT
CC SIGNALING PATHWAY FROM RECEPTORS TO THE NUCLEUS.
CC -1- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
CC PROTEIN KINASES. BELONGS TO THE ML/RAF SUBFAMILY.
CC -1- SIMILARITY: CONTAINS A COPY OF THE ZINC-DEPENDENT PHOSBO-
CC ESTER AND DAG BINDING DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: X12948; G65028; -
 CC EMBL: S74063; G241260; -
 CC PIR: S01930; TVXLRP.

DR PROSITE: P500107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: P500108; PROTEIN_KINASE_ST; 1.
 DR PROSITE: P500479; DAG_PE_BINDING_DOMAIN; 1.
 DR PROSITE: P550011; PROTEIN_KINASE_DOM; 1.
 DR PFAM: PF00069; PKinase; 1.
 DR PFAM: PF00130; DAG-pe-bind; 1.
 DR HSP: P04049; 1FQO.
 DR TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; PROTO-ONCOGENE; ZINC;
 DR ATP-BINDING; PHORBOL-ESTER BINDING.
 DR DOMAIN 138 183 PHORBOL-ESTER AND DAG BINDING.
 DR NP_BIND 340 600 PROTEIN KINASE.
 DR BINDING 346 354 ATP (BY SIMILARITY).
 DR ACT_SITE 356 366 ATP (BY SIMILARITY).
 DR CONFLICT 459 459 BY SIMILARITY.
 DR CONFLICT 309 309 K -> R (IN REF. 2).
 SQ SEQUENCE 638 AA; 71959 MW; A46048CC CRC32;

Query Match 43.28; Score 64; DB 1; Length 638;
 Best Local Similarity 30.08; Pred. No. 1.94e+00;
 Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

DB 404 TKDNLAVTQWCGSSLYH 423
 1 ANDHSLLEAMSDNTPLYH 20

RESULT 13
 ID PMAL DUNBI STANDARD; PRT; 1131 AA.
 AC P54211;
 DT 01-OCT-1996 (REL. 34, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE PLASMA MEMBRANE ATPASE (EC 3.6.1.35) (PROTON PUMP).
 GN PMAL.
 OS DUNALIELLA BIOCULATA.
 OS EUKARYOTA; VIRIDIPHYTES; CHLOROPHYTA; CHLOROPHYCEAE; VOLVOCALES;
 OC DUNALIACEAE; DUNALIELLA.
 [1]
 SEQUENCE FROM N.A.
 STRAIN-SAG 19-4;
 MEDLINE; 95375230.
 RA WOLF A.H., SLATMAN C.W., GRADMAN D.;
 RT "Primary structure of the plasma membrane H(+)-ATPase from the
 RT halotolerant alga Dunaliella bioculata.";
 RL PLANT MOL. BIOL. 28:657-666(1995).
 CC -1- FUNCTION: THE PLASMA MEMBRANE ATPASE OF PLANTS AND FUNGI IS A
 CC HYDROGEN ION PUMP. THE PROTON GRADIENT IT GENERATES DRIVES THE
 CC ACTIVE TRANSPORT OF NUTRIENTS BY H+-SYMPORT. THE RESULTING
 CC EXTERNAL ACIDIFICATION AND/OR INTERNAL ALKALIZATION MAY MEDIATE
 CC GROWTH RESPONSES.
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN
 CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
 CC (E1-E2 ATPASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC EMBL: X73901; G313704; -
 CC PROSITE: P500154; ATPASE_E1_E2; 1.
 DR PFAM: PF00122; E1-E2 ATPase; 1.
 KM HYDROLASE; HYDROGEN ION TRANSPORT; TRANSMEMBRANE; PHOSPHORYLATION;
 KM ATP-BINDING.
 FT TRANSFER 77 97 POTENTIAL.
 FT TRANSFER 98 118 POTENTIAL.
 FT TRANSFER 151 171 POTENTIAL.
 FT TRANSFER 231 251 POTENTIAL.
 FT TRANSFER 265 285 POTENTIAL.
 FT TRANSFER 305 325 POTENTIAL.
 FT TRANSFER 642 662 POTENTIAL.
 FT TRANSFER 689 709 POTENTIAL.
 FT TRANSFER 733 753 POTENTIAL.
 FT TRANSFER 884 904 POTENTIAL.
 FT TRANSFER 946 966 POTENTIAL.
 FT MOD_RES 357 357 PHOSPHORYLATION (PROBABLE).
 SQ SEQUENCE 1131 AA; 123450 MW; A71FE561 CRC32;

Query Match 42.6%; Score 63; DB 1; Length 1131;
 Best Local Similarity 35.7%; Pred. No. 2.87e+00;
 Matches 5; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

DB 774 NDQSYVLPQPCDE 787
 2 NDHSLLEAMSDND 15

RESULT 14
 ID KRAF MSV36 STANDARD; PRT; 323 AA.
 AC P00532;
 DT 21-JUL-1986 (REL. 01, CREATED)
 DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
 DE RAF SERINE/THREONINE-PROTEIN KINASE TRANSFORMING PROTEIN (EC 2.7.1.-).
 GN V-RAF.
 OS MURINE SARCOMA VIRUS 3611.
 OS VIRUSES; RETROID VIRUSES; RETROVIRIDAE; MAMMALIAN TYPE C RETROVIRUSES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RP MEDLINE; 84121298.
 RA KAN N.C., FLORELLIS C.S., MARK G.E., DUESBERG P.H., PAPAS T.S.;
 RT "A common onc gene sequence transduced by avian carcinoma virus MH2
 RT and by murine sarcoma virus 3611.";
 RL SCIENCE 223:813-816(1984).
 RN [2]
 RP SEQUENCE FROM N.A.
 RP MEDLINE; 84172180.
 RA MARK G.E., RAPP U.R.;
 RT "Primary structure of v-raf: relatedness to the src family of
 RT oncogenes.";
 RL SCIENCE 224:285-289(1984).
 CC -1- THIS PROTEIN IS SYNTHESIZED AS A GAG-RAF POLYPEPTIDE.
 CC -1- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SRP/THR-
 CC PROTEIN KINASES. BELONGS TO THE ML/RAF SUBFAMILY
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: K01691; G332207; ALT_INIT.
 DR PIR: A00638; TVWVP6.
 DR PROSITE: P500107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: P500108; PROTEIN_KINASE_ST; 1.
 DR PROSITE: P550011; PROTEIN_KINASE_DOM; 1.
 DR PFAM: PF00069; PKinase; 1.
 DR HSP: P11362; 1FGI.

KW POLYPROTEIN: TRANSFORMING PROTEIN; SERINE/THREONINE-PROTEIN KINASE;
 KM TRANSFERASE: ONCOGENE: ATP-BINDING. PROTEIN KINASE.
 FT DOMAIN 24 284
 FT NP_BIND 30 38 ATP (BY SIMILARITY).
 FT BINDING 50 50 ATP (BY SIMILARITY).
 FT ACT SITE 143 143 BY SIMILARITY.
 SQ SEQUENCE 323 AA: 36883 MW: 8355486 CRC32:

Query Match 41.9%; Score 62; DB 1; Length 323;
 Best Local Similarity 30.0%; Pred. No. 4.23e+00;
 Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

Db 88 TKDNLAIYVQWCGSSLYKH 107
 OY 1 ANDHLSLEAMSDNDPYLH 20

RESULT 15
 KMIL_AVTMH STANDARD; PRT; 380 AA.

DI 21-JUL-1986 (REL. 01, CREATED)
 DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE MIL SERINE/THREONINE-PROTEIN KINASE TRANSFORMING PROTEIN
 DE (EC 2.7.1.1).
 GN V-MIL OR V-MHT.
 OS AVIAN RETROVIRUS MH2.
 OC VIRUSES; RETROVIRIDAE; AVIAN TYPE C RETROVIRUSES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 84191511.
 RA SOTRAVE P., BONNER T.I., RAPP U.R., JANSEN H.W., PATSCHINSKY T.,
 RA BISTER K.;
 RT "Nucleotide sequence of avian retroviral oncogene v-mil: homologue of
 RT murine retroviral oncogene v-raf.";
 RL NATURE 309:85-88(1984).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 84121298.
 RA KAN N.C., FLOREDELIS C.S., MARK G.E., DUESBERG P.H., PAPAS T.S.;
 RT "A common onc gene sequence transduced by avian carcinoma virus MH2
 RT and by murine sarcoma virus 3611.";
 RL SCIENCE 223:813-816(1984).
 CC -1- DISEASE: BY ITSELF THE V-MIL ONCOGENE HAS ONLY WEAK TRANSFORMING
 CC CAPACITY BUT IT ABOLISHES THE GROWTH FACTOR REQUIREMENTS OF AVIAN
 CC MACROPHAGES TRANSFORMED BY OTHER ONCOGENES.
 CC -1- THIS PROTEIN IS SYNTHESIZED AS A GAG-MHT OR GAG-MIL POLYPROTEIN.
 CC -1- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
 CC PROTEIN KINASES. BELONGS TO THE MIL/RAF SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X00534; E235847; ALT_INIT.
 DR PIR: A00639; TVFVNM.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PFAM: PF00069; Pkinase; 1.
 DR HSSP: P11362; IFGI.
 KM POLYPROTEIN: TRANSFORMING PROTEIN; SERINE/THREONINE-PROTEIN KINASE;
 KM TRANSFERASE: ONCOGENE: ATP-BINDING. PROTEIN KINASE.
 FT DOMAIN 82 341
 FT NP_BIND 88 96 ATP (BY SIMILARITY).
 FT BINDING 108 108 ATP (BY SIMILARITY).
 FT ACT SITE 201 201 BY SIMILARITY.
 FT CONFLICT 211 211 G -> E (IN REF. 2).
 SQ SEQUENCE 380 AA: 42853 MW: 1078284 CRC32:

Query Match 41.9%; Score 62; DB 1; Length 380;
 Best Local Similarity 30.0%; Pred. No. 4.23e+00;
 Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

Db 146 TKDNLAIYVQWCGSSLYKH 165
 OY 1 ANDHLSLEAMSDNDPYLH 20

Search completed: Tue Jan 11 15:37:35 2000
 Job time : 9 secs.

 NWSETH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MSrch_PP protein - protein database search, using Smith-Waterman algorithm
 Run on: Tue Jan 11 15:47:05 2000; MasPar time 3.48 Seconds
 253.134 Million cell updates/sec
 No output not generated.

Title: >US-09-290-049-3
 Description: (1-22) from US09290049.pep
 Perfect Score: 155
 Sequence: 1 TGARTINGOLLYFRANGVQVKG 22

Scoring table: PAM 150
 Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: p1r60
 1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 29.032; Variance 42.944; scale 0.676

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	155	100.0	1475	2 B33135	gtfB protein precursor	4.92e-20
2	112	72.3	1592	2 A38175	glucosyltransferase p	5.30e-10
3	109	70.3	1518	2 A44811	glucosyltransferase (2.45e-09
4	107	69.0	1290	2 JC5473	dextranucrase (EC 2.	6.75e-09
5	106	68.4	1375	2 UT0345	dextranucrase (EC 2.	1.12e-08
6	106	68.4	1375	2 S22737	glucosyltransferase (1.12e-08
7	104	67.1	1365	2 A41483	glucosyltransferase (3.05e-08
8	93	60.0	1431	2 A45866	dextranucrase (EC 2.	6.72e-06
9	88	56.8	1431	2 A37184	glucan-binding protei	7.21e-05
10	83	53.5	329	2 A55221	dextranase inhibitor p	7.32e-04
11	79	51.0	382	2 S55166	hypothetical protein	4.46e-03
12	77	49.7	329	2 F71014	probable dipeptidase tr	1.08e-02
13	72	46.5	126	2 A40016	doc protein - phase p	9.45e-02
14	71	45.8	425	2 H70410	UDP-N-acetylglucosami	1.44e-01
15	68	43.9	831	3 T00323	chitinase (EC 3.2.1.1	5.03e-01
16	67	43.2	245	2 F64465	hypothetical protein	7.58e-01
17	67	43.2	2710	2 A37052	toxigenin A - Clostridi	7.58e-01
18	65	41.9	822	2 B54846	fibroblast growth fac	1.70e+00
19	65	41.9	1396	2 S36851	L-shaped tail fiber p	1.70e+00
20	63	40.6	240	2 H69366	conserved hypothetical	3.75e+00
21	63	40.6	301	2 S13165	astalogycoprotein re	3.75e+00
22	63	40.6	310	2 F71020	probable fructokinase	3.75e+00
23	63	40.6	396	1 WZBE8	dUTP pyrophosphatase	3.75e+00

ALIGNMENTS

RESULT ENTRY	1	B33135	#type complete	2	A54871	Gal beta-1, 3galNAc-s	3.75e+00
TITLE	gtfB protein precursor	-	Streptococcus mutans	2	S65290	chitin-associated p	3.75e+00
ORGANISM	Streptococcus mutans			2	S74633	high affinity sulfate	3.75e+00
DATE	23-Oct-1990	#sequence	revision 23-Oct-1990	2	1082	inner layer protein V	3.75e+00
ACCESSIONS	09-Sep-1997	#sequence	revision 23-Oct-1990	2	28	L2 protein - human pa	5.53e+00
REFERENCE	B33135; A33128			2	29	L2 protein - human pa	5.53e+00
#authors	A33135			2	30	isfp protein - f1ssio	5.53e+00
#journal	Shiroza, T.; Ueda, S.; Kuramitsu, H.K.			2	31	L2 protein - human pa	5.53e+00
#cross-references	J. Bacteriol. (1987) 169:4263-4270			2	32	cytochrome L - Clostri	5.53e+00
#accession	Sequence analysis of the gtfB gene from Streptococcus mutans.			2	33	uroporphyrin-III c-me	8.12e+00
#status	B33135			2	34	int protein - phage 1	8.12e+00
#molecule-type	preliminary			2	35	hypothetical protein	8.12e+00
#residues	1-1475	#label SHI		2	36	hypothetical protein	8.12e+00
#cross-references	GB:M17361; NID:gl53639; PID:gl53640			2	37	ABC-type transport pr	8.12e+00
REFERENCE	A33128			2	38	cannabinoid receptor	1.19e+01
#authors	Shiroza, T.; Ueda, S.; Kuramitsu, H.K.			2	39	hypothetical protein	1.19e+01
#journal	J. Bacteriol. (1987) 169:4263-4270			2	40	hypothetical protein	1.19e+01
#cross-references	Sequence analysis of the gtfB gene from Streptococcus mutans.			2	41	enterotoxin A - Clost	1.19e+01
#accession	B33135			2	42	transcription factor	1.19e+01
#status	preliminary			2	43	gene try protein - ra	1.19e+01
#molecule-type	1-1475	#label SHI		2	44	regulatory protein Sp	1.19e+01
#residues	1-1475	#label SHI		2	45	Spl protein - rat	1.19e+01
#cross-references	GB:M17361; NID:gl53639; PID:gl53640			2			
REFERENCE	A33128			2			
#authors	Shiroza, T.; Ueda, S.; Kuramitsu, H.K.			2			
#journal	J. Bacteriol. (1987) 169:4263-4270			2			
#cross-references	Sequence analysis of the gtfB gene from Streptococcus mutans.			2			
#accession	B33135			2			
#status	preliminary			2			
#molecule-type	1-1475	#label SHI		2			
#residues	1-1475	#label SHI		2			
#cross-references	GB:M17361; NID:gl53639; PID:gl53640			2			
REFERENCE	A33128			2			
#authors	Shiroza, T.; Ueda, S.; Kuramitsu, H.K.			2			
#journal	J. Bacteriol. (1987) 169:4263-4270			2			
#cross-references	Sequence analysis of the gtfB gene from Streptococcus mutans.			2			
#accession	B33135			2			
#status	preliminary			2			
#molecule-type	1-1475	#label SHI		2			
#residues	1-1475	#label SHI		2			
#cross-references	GB:M17361; NID:gl53639; PID:gl53640			2			
REFERENCE	A33128			2			
#authors	Shiroza, T.; Ueda, S.; Kuramitsu, H.K.			2			
#journal	J. Bacteriol. (1987) 169:4263-4270			2			
#cross-references	Sequence analysis of the gtfB gene from Streptococcus mutans.			2			
#accession	B33135			2			
#status	preliminary			2			
#molecule-type	1-1475	#label SHI		2			
#residues	1-1475	#label SHI		2			
#cross-references	GB:M17361; NID:gl53639; PID:gl53640			2			
REFERENCE	A33128			2			
#authors	Shiroza, T.; Ueda, S.; Kuramitsu, H.K.			2			
#journal	J. Bacteriol. (1987) 169:4263-4270			2			
#cross-references	Sequence analysis of the gtfB gene from Streptococcus mutans.			2			
#accession	B33135			2			
#status	preliminary			2			
#molecule-type	1-1475	#label SHI		2			
#residues	1-1475	#label SHI		2			
#cross-references	GB:M17361; NID:gl53639; PID:gl53640			2			
REFERENCE	A33128			2			
#authors	Shiroza, T.; Ueda, S.; Kuramitsu, H.K.			2			
#journal	J. Bacteriol. (1987) 169:4263-4270			2			
#cross-references	Sequence analysis of the gtfB gene from Streptococcus mutans.			2			
#accession	B33135			2			
#status	preliminary			2			
#molecule-type	1-1475	#label SHI		2			
#residues	1-1475	#label SHI		2			
#cross-references	GB:M17361; NID:gl53639; PID:gl53640			2			
REFERENCE	A33128			2			
#authors	Shiroza, T.; Ueda, S.; Kuramitsu, H.K.			2			
#journal	J. Bacteriol. (1987) 169:4263-4270			2			
#cross-references	Sequence analysis of the gtfB gene from Streptococcus mutans.			2			
#accession	B33135			2			
#status	preliminary			2			
#molecule-type	1-1475	#label SHI		2			
#residues	1-1475	#label SHI		2			
#cross-references	GB:M17361; NID:gl53639; PID:gl53640			2			
REFERENCE	A33128			2			
#authors	Shiroza, T.; Ueda, S.; Kuramitsu, H.K.			2			
#journal	J. Bacteriol. (1987) 169:4263-4270			2			
#cross-references	Sequence analysis of the gtfB gene from Streptococcus mutans.			2			
#accession	B33135			2			
#status	preliminary			2			
#molecule-type	1-1475	#label SHI		2			
#residues	1-1475	#label SHI		2			
#cross-references	GB:M17361; NID:gl53639; PID:gl53640			2			
REFERENCE	A33128			2			
#authors	Shiroza, T.; Ueda, S.; Kuramitsu, H.K.			2			
#journal	J. Bacteriol. (1987) 169:4263-4270			2			
#cross-references	Sequence analysis of the gtfB gene from Streptococcus mutans.			2			
#accession	B33135			2			
#status	preliminary			2			
#molecule-type	1-1475	#label SHI		2			
#residues	1-1475	#label SHI		2			
#cross-references	GB:M17361; NID:gl53639; PID:gl53640			2			
REFERENCE	A33128			2			
#authors	Shiroza, T.; Ueda, S.; Kuramitsu, H.K.			2			
#journal	J. Bacteriol. (1987) 169:4263-4270			2			
#cross-references	Sequence analysis of the gtfB gene from Streptococcus mutans.			2			
#accession	B33135			2			
#status	preliminary			2			
#molecule-type	1-1475	#label SHI		2			
#residues	1-1475	#label SHI		2			
#cross-references	GB:M17361; NID:gl53639; PID:gl53640			2			
REFERENCE	A33128			2			
#authors	Shiroza, T.; Ueda, S.; Kuramitsu, H.K.			2			
#journal	J. Bacteriol. (1987) 169:4263-4270			2			
#cross-references	Sequence analysis of the gtfB gene from Streptococcus mutans.			2			
#accession	B33135			2			
#status	preliminary			2			
#molecule-type	1-1475	#label SHI		2			
#residues	1-1475	#label SHI		2			
#cross-references	GB:M17361; NID:gl53639; PID:gl53640			2			
REFERENCE	A33128			2			
#authors	Shiroza, T.; Ueda, S.; Kuramitsu, H.K.			2			
#journal	J. Bacteriol. (1987) 169:4263-4270			2			
#cross-references	Sequence analysis of the gtfB gene from Streptococcus mutans.			2			
#accession	B33135			2			
#status	preliminary			2			
#molecule-type	1-1475	#label SHI		2			
#residues	1-1475	#label SHI		2			
#cross-references	GB:M17361; NID:gl53639; PID:gl53640			2			
REFERENCE	A33128			2			
#authors	Shiroza, T.; Ueda, S.; Kuramitsu, H.K.			2			
#journal	J. Bacteriol. (1987) 169:4263-4270			2			
#cross-references	Sequence analysis of the gtfB gene from Streptococcus mutans.			2			
#accession	B33135			2			
#status	preliminary			2			
#molecule-type	1-1475	#label SHI		2			
#residues	1-1475	#label SHI		2			
#cross-references	GB:M17361; NID:gl53639; PID:gl53640			2			
REFERENCE	A33128			2			
#authors	Shiroza, T.; Ueda, S.; Kuramitsu, H.K.			2			
#journal	J. Bacteriol. (1987) 169:4263-4270			2			
#cross-references	Sequence analysis of the gtfB gene from Streptococcus mutans.			2			
#accession	B33135			2			
#status	preliminary			2			
#molecule-type	1-1475	#label SHI		2			
#residues	1-1475	#label SHI		2			
#cross-references	GB:M17361; NID:gl53639; PID:gl53640			2			
REFERENCE	A33128			2			
#authors	Shiroza, T.; Ueda, S.; Kuramitsu, H.K.			2			
#journal	J. Bacteriol. (1987) 169:4263-4270			2			
#cross-references	Sequence analysis of the gtfB gene from Streptococcus mutans.			2			
#accession	B33135			2			
#status	preliminary			2			
#molecule-type	1-1475	#label SHI		2			
#residues	1-1475	#label SHI		2			
#cross-references	GB:M17361; NID:gl53639; PID:gl53640			2			
REFERENCE	A33128			2			
#authors	Shiroza, T.; Ueda, S.; Kuramitsu, H.K.			2			
#journal	J. Bacteriol. (1987) 169:4263-4270			2			
#cross-references	Sequence analysis of the gtfB gene from Streptococcus mutans.			2			
#accession	B33135			2			
#status	preliminary			2			
#molecule-type	1-1475	#label SHI		2			
#residues	1-1475	#label SHI		2			
#cross-references	GB:M17361; NID:gl53639; PID:gl53640			2			
REFERENCE	A33128			2			
#authors	Shiroza, T.; Ueda, S.; Kuramitsu, H.K.			2			
#journal	J. Bacteriol. (1987) 169:4263-4270			2			
#cross-references	Sequence analysis of the gtfB gene from Streptococcus mutans.			2			
#accession	B33135			2			
#status	preliminary			2			
#molecule-type	1-1475	#label SHI</					

DATE 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change 09-Sep-1997

ACCESSIONS A38175

ENTRY A38175

TITLE Abo, H.; Mtsunura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.

ORGANISM J. Bacteriol. (1991) 173:989-996

REFERENCE Peptide sequences for sucrose splitting and glucan binding within Streptococcus sobrius glucosyltransferase (water-insoluble glucan synthetase).

FEATURES

##cross-references MUID:91123227

##accession A38175

##status Preliminary

##molecule_type DNA

##residues 1-1592 ##label ABO

##cross-references GB:D90213; NID:9217032; PID:41014946; PID:9217033

CLASSIFICATION #superfamily cpl repeat homology

FEATURE

1093-1112 #domain cpl repeat homology #label CP1\

1222-1241 #domain cpl repeat homology #label CP2\

1287-1306 #domain cpl repeat homology #label CP3\

1330-1351 #domain cpl repeat homology #label CP4\

1352-1371 #domain cpl repeat homology #label CP5\

1402-1420 #domain cpl repeat homology #label CP6\

1465-1484 #domain cpl repeat homology #label CP7\

1513-1532 #domain cpl repeat homology #label CP8\

SUMMARY #length 1592 #molecular-weight 176167 #checksum 5940

Query Match 72.3%; Score 112; DB 2; Length 1592;

Best Local Similarity 77.3%; Pred. No. 5.30e-10;

Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 1298 TGAATGKQKLYFRANGQVKG 1319

QY 1 TGARTINGQLLYFRANGQVKG 22

RESULT 3

ENTRY A44811

TITLE #type complete

ORGANISM glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius

REFERENCE #formal_name Streptococcus salivarius

DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Sep-1997

ACCESSIONS A44811; S22726; S28809

REFERENCE A44811

ENTRY Giffard, P.M.; Simpson, C.L.; Milyard, C.P.; Jacques, N.A.

TITLE J. Gen. Microbiol. (1991) 137:2577-2593

ORGANISM Molecular characterization of a cluster of at least two glucosyltransferase genes in Streptococcus salivarius ATCC 25975.

FEATURES

##cross-references MUID:92148377

##accession A44811

##molecule_type DNA

##residues 1-1518 ##label GIF

##cross-references EMBL:211873; NID:947526; PID:947527

##note Sequence extracted from NCBI backbone (NCBIN:81050, NCBIPI:81052)

GENETICS

##gene gtff

CLASSIFICATION #superfamily cpl repeat homology

KEYWORDS glycosyltransferase; hexosyltransferase

FEATURE

1307-1326 #domain cpl repeat homology #label CP4

SUMMARY #length 1518 #molecular-weight 167730 #checksum 1334

Query Match 70.3%; Score 109; DB 2; Length 1518;

Best Local Similarity 72.7%; Pred. No. 2.45e-09;

Matches 16; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Db 1383 TGAQVINGOHLYFNADSGQVKG 1404

QY 1 TGARTINGQLLYFRANGQVKG 22

RESULT 4

ENTRY JC5473

TITLE #type complete

ORGANISM dextranase (EC 2.4.1.5) - Leuconostoc mesenteroides

REFERENCE #formal_name Leuconostoc mesenteroides

DATE 07-Jul-1997 #sequence_revision 29-Aug-1997 #text_change 29-Aug-1997

ACCESSIONS JC5473

ENTRY JC5473

TITLE Monahan, P.

ORGANISM Gene (1996) 182:23-32

REFERENCE Cloning and sequencing of a gene coding for a novel dextranase from Leuconostoc mesenteroides NRRL B-1299 synthesizing only alpha (1-6) and alpha (1-3) linkages.

FEATURES

##cross-references MUID:97136686

##accession JC5473

##status nucleic acid sequence not shown; translation not shown

##molecule_type DNA

##residues 1-1290 ##label MON

##cross-references GB:U38181

COMMENT This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose onto acceptor molecules.

GENETICS

##gene dsra

KEYWORDS glycosyltransferase; hexosyltransferase

FEATURE

78-870 #domain catalytic #status predicted #label CAT\

922-1290 #domain glucan-binding #status predicted #label GCB

SUMMARY #length 1290 #molecular-weight 145572 #checksum 9386

Query Match 69.0%; Score 107; DB 2; Length 1290;

Best Local Similarity 63.6%; Pred. No. 6.75e-09;

Matches 14; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 1181 TGDRIIRKQNLRYNONGKMG 1202

QY 1 TGARTINGQLLYFRANGQVKG 22

RESULT 5

ENTRY JT0345

TITLE #type complete

ORGANISM dextranase (EC 2.4.1.5) precursor - Streptococcus mutans

REFERENCE #formal_name Streptococcus mutans

DATE 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 22-Nov-1996

ACCESSIONS JT0345; C33135

REFERENCE JT0345

ENTRY Ueda, S.; Shiroya, T.; Kuramitsu, H.K.

TITLE Gene (1988) 69:101-109

ORGANISM Sequence analysis of the gtfc gene from Streptococcus mutans GS-5

FEATURES

##cross-references MUID:89137980

##accession JT0345

##molecule_type DNA

##residues 1-1375 ##label UED

##experimental_source GS-5

REFERENCE A33135

ENTRY Shiroya, T.; Ueda, S.; Kuramitsu, H.K.

TITLE J. Bacteriol. (1987) 169:4263-4270

ORGANISM Sequence analysis of the gtfc gene from Streptococcus mutans.

FEATURES

##cross-references MUID:8730803

##accession C33135

##status Preliminary

##molecule_type DNA

##residues 1-349 ##label SHI

##cross-references GB:M17361

GENETICS

##gene gtfc

FUNCTION #description catalyzes the synthesis of both water-soluble and

#authors Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.

*authors Banas, J.A.; Russell, R.R.B.; Ferretti, J.J.
#journal Infect. Immun. (1990) 58:667-673

```

#title      Sequence analysis of the gene for the glucan-binding protein
             of Streptococcus mutans Inspritt.
#cross-references MUID:90170123
#accession   A37184
#status      preliminary
#molecule_type DNA
#residues    1563 #label BAN
#cross-references GB:M50945; MID:g153637; PID:g153638
CLASSIFICATION #superfamily cpl repeat homology
FEATURE
169-188      #domain cpl repeat homology #label CP1\
264-283      #domain cpl repeat homology #label CP2\
349-368      #domain cpl repeat homology #label CP3\
504-523      #domain cpl repeat homology #label CP4\
525-548      #domain cpl repeat homology #label CP5
SUMMARY
#length 563 #molecular-weight 62932 #checksum 2435

Query Match      56.88; Score 88; DB 2; Length 563;
                  St. Local Similarity 68.48; Pred. No. 7.21e-05;
                  Lches 13; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Db      278 RTIGKKRYEDTNGQYKG 296
      ||| | | | | | | | | |
OY      4 RTINGQLLYFRANGQYKG 22

RESULT 10
ENTRY   A55221 #type complete
TITLE   dextranase inhibitor precursor - Streptococcus sobrinus
         (strain UA866, serotype g)
ORGANISM #formal_name Streptococcus sobrinus
          23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change
          17-Mar-1999
ACCESSIONS A55221
REFERENCE   Sun, J.W.; Wanda, S.Y.; Camilli, A.; Curtiss III, R.
#authors    J. Bacteriol. (1994) 176:7213-7222
#journal     Cloning and DNA sequencing of the dextranase inhibitor gene
#title       (del) from Streptococcus sobrinus.
#cross-references MUID:95050304
#accession   A55221
#status      preliminary
#molecule_type DNA
#residues    1-329 #label SUN
#cross-references GB:L34406
CLASSIFICATION #superfamily cpl repeat homology
FEATURE
173-192      #domain cpl repeat homology #label CP1
          #length 329 #molecular-weight 36077 #checksum 8218

Query Match      53.58; Score 83; DB 2; Length 329;
                  Best Local Similarity 68.28; Pred. No. 7.32e-04;
                  Matches 15; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

Db      185 GLGTGQTLFEDNTGQYKG 206
      | : | | | | | | | | | |
OY      2 GARTINGQLLYFRAN-GVQYKG 22

RESULT 11
ENTRY   S55166 #type complete
TITLE   hypothetical protein YOL147c - yeast (Saccharomyces
         cerevisiae)
ALTERNATE_NAMES #formal_name Saccharomyces cerevisiae
ORGANISM        08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change
DATE            05-Jun-1998
ACCESSIONS      S55166; S56929; S71650
REFERENCE       Katsoulou, C.; Tzermia, M.; Alexandraki, D.
#authors        submitted to the EMBL Data Library, May 1995
#submission     The complete sequence of a 40.7 kb segment located on the
#description     left arm of yeast chromosome X identified 13 known genes

```

```

#accession   S55166
#molecule_type DNA
#residues    1-382 #label KAT
#cross-references EMBL:X87371; MID:9854542; PID:9854550
REFERENCE      S56912
#authors       Katsoulou, C.; Tzermia, M.; Alexandraki, D.
#submission    submitted to the Protein Sequence Database, September 1995
#accession     S56929
#molecule_type DNA
#residues     1-382 #label KAM
#cross-references EMBL:Z49422; MID:g1015560; PID:g1015561; MIPS:YOL147c
REFERENCE      S71643
#authors       Katsoulou, C.; Tzermia, M.; Tavernarakis, N.; Alexandraki, D.
#journal       Yeast (1996) 12:787-797
#title         Sequence analysis of a 40.7 kb segment from the left arm of
                  yeast chromosome X reveals 14 known genes and 13 new open
                  reading frames including homologues of genes clustered on
                  the right arm of chromosome XI.
#accession     S71650
#status        nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues     1-382 #label KAF
#cross-references EMBL:X87371; MID:9854542; PID:9854550
#note          the nucleotide sequence was submitted to the EMBL Data
                  Library, May 1995

GENETICS
#map_position 101.
SUMMARY #length 382 #molecular-weight 44862 #checksum 8833

Query Match      51.08; Score 79; DB 2; Length 382;
                  Best Local Similarity 38.98; Pred. No. 4.46e-03;
                  Matches 7; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

Db      170 RYVISTILRYRTDVOIR 187
      | : | | | | | | | | | |
OY      4 RTINGQLLYFRANGQYKG 21

RESULT 12
ENTRY   F71014 #type complete
TITLE   probable dipeptide transport ATP-binding protein dppd -
         Pyrococcus horikoshii
ORGANISM #formal_name Pyrococcus horikoshii
          14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change
          21-Nov-1998
ACCESSIONS      F71014
REFERENCE       Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.;
#authors        Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;
                  Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;
                  Nakazawa, H.; Takamiya, M.; Ohitoku, Y.; Funahashi, T.;
                  Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kusuda, N.; Oguchi,
                  A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;
                  Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
#journal        DNA Res. (1998) 5:55-76
#title          Complete sequence and gene organization of the genome of a
                  hyper-thermophilic archaeobacterium, Pyrococcus horikoshii
                  OT3.
#cross-references MUID:98344137
#accession     F71014
#status        preliminary; nucleic acid sequence not shown;
                  translation not shown
#molecule_type DNA
#residues     1-329 #label KAM
#cross-references GB:AP000006; MID:g3236133; PID:d1031461; PID:g3257835
#experimental_source strain OT3
#note          this accession replaces an interim accession for a
                  sequence replaced by Genbank

GENETICS
#note          PH1412
FUNCTION

```

```

#description probably responsible for energy-coupling to the transport
#system
CLASSIFICATION #superfamily inner membrane protein malK; ATP-binding
#accession H70410
KEYWORDS cassette homology
AMP; binding protein-dependent transport system; nucleotide
binding; oligopeptide transport; P-loop
FEATURE
28-241 #domain ATP-binding cassette homology #label ABC\
45-53 #region nucleotide-binding motif A (P-loop)\
185-189 #region nucleotide-binding motif B
SUMMARY #length 329 #molecular-weight 36504 #checksum 6349

Query Match 49.78; Score 77; DB 2; Length 329;
Best Local Similarity 44.48; Pred. No. 1.08e-02;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Db 66 GRVGGVLYRREDGSV 83
QY 3 ARTINGQLYFRANGVOV 20

#cross-references M01D:94016561
#accession S40016 #type complete
#molecule_type DNA
#residues 1-126 #label LEH
#cross-references GB:M95666; NID:g463276; PID:g463277
REFERENCE S38553
#authors Schmidt, C.; Lehnher, H.; Guldolin, A.; Aber, W.
#submission Submitted to the EMBL Data Library, November 1992
#description Additional late promoter sequences of bacteriophage P1.
#accession S38555
#molecule_type DNA
#residues 26-126 #label SCH
#cross-references EMBL:M95666
SUMMARY #length 126 #molecular-weight 13588 #checksum 6874

Query Match 46.58; Score 72; DB 2; Length 126;
Best Local Similarity 56.38; Pred. No. 9.45e-02;
Matches 9; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 76 ALNSALLFRNGVOV 91
QY 5 TINGQLYFRANGVOV 20

RESULT 14
ENTRY H70410 #type complete
TITLE UDP-N-acetylglucosamine 1-carboxyvinyltransferase - Aquifex
aeolicus
ORGANISM #formal_name Aquifex aeolicus
DATE 08-May-1998 #sequence_revision 08-May-1998 #text_change
13-Sep-1998
ACCESSIONS H70410
REFERENCE A70300
#authors Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;
Lenox, A.L.; Graham, D.E.; Overbeek, R.; Sneed, M.A.;
Keller, M.; Aulay, M.; Huber, R.; Feldman, R.A.; Short,
J.M.; Olson, G.J.; Swanson, R.V.
#journal Nature (1998) 392:353-358
#title The complete genome of the hyperthermophilic bacterium

```

```

#cross-references M01D:98196666
#accession H70410
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-425 #label AOF
#cross-references GB:AE000732; NID:g2983704; PID:g2983705; GB:AE000657
#experimental_source strain VF5
GENETICS
#gene mura
CLASSIFICATION #superfamily UDP-N-acetylglucosamine
1-carboxyvinyltransferase MurZ
SUMMARY #length 425 #molecular-weight 47259 #checksum 7026

Query Match 45.88; Score 71; DB 2; Length 425;
Best Local Similarity 42.18; Pred. No. 1.44e-01;
Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Db 127 GARPIDHLEKFEKAGADV 145
QY 2 GARTINGQLYFRANGVOV 20

RESULT 15
ENTRY T00323 #type complete
TITLE chitinase (EC 3.2.1.14) B - Clostridium paraputrificum
ORGANISM #formal_name Clostridium paraputrificum
DATE 29-Jan-1999 #sequence_revision 29-Jan-1999 #text_change
T00323
ACCESSIONS T00323
REFERENCE 214138
#authors Ohmiva, K.; Sakka, K.; Kimura, T.; Karita, S.; Morimoto, K.
#journal J. Bacteriol. (1997) 179:7306-7314
#title Cloning, sequencing, and expression of the gene encoding
Clostridium paraputrificum chitinase ChA and analysis of
the functions of novel cadherin-like domains and a
chitin-binding domain.
#accession T00323
#status preliminary
#residues 1-831 #label OHM
#cross-references EMBL:AB001874; NID:d1175645; PID:d1024701
SUMMARY #length 831 #molecular-weight 90021 #checksum 7845

Query Match 43.98; Score 68; DB 3; Length 831;
Best Local Similarity 36.48; Pred. No. 5.03e-01;
Matches 8; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

Db 525 TAATYKNNKVIYIKNGVAITG 546
QY 1 TGARTINGQLYFRANGVOVG 22

Search completed: Tue Jan 11 15:47:33 2000
Job time : 28 secs.

```

 W I D E S E E T H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mparch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Jan 11 15:36:55 2000; Maspar time 3.63 Seconds
 232.086 Million cell updates/sec

 No output not generated.

Title: >US-09-290-049-1
 Description: (1-21) from US09290049.pep
 Perfect Score: 148
 Sequence: 1 ANDHLSILEAWSNDPYLHD 21

Scoring table: PAM 150
 Gap 15

Searched: 122810 segs, 40068593 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: p160
 1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 28.783; Variance 46.009; scale 0.626

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	148	100.0	1475	2	B33135	7.09e-17
2	136	91.9	1375	2	UT0345	3.13e-14
3	131	88.5	1592	2	A38175	3.81e-13
4	102	68.9	1365	2	A41483	4.40e-07
5	101	68.2	1431	2	A45866	6.99e-07
6	78	52.7	1599	2	S22737	1.70e-02
7	75	50.7	1599	2	A44811	5.82e-02
8	72	48.6	131	2	B48445	1.94e-01
9	72	48.6	336	1	Q08E40	1.94e-01
10	72	48.6	579	2	E71557	1.94e-01
11	67	45.3	490	2	H70538	1.36e+00
12	66	44.6	1885	2	JC4086	1.98e+00
13	64	43.2	476	2	S58229	4.19e+00
14	64	43.2	480	1	A30065	4.19e+00
15	64	43.2	515	2	E71447	4.19e+00
16	64	43.2	542	2	T02301	4.19e+00
17	64	43.2	638	1	TVXLRP	4.19e+00
18	64	42.6	1131	2	S34213	6.06e+00
19	62	41.9	323	1	TVFVMS	8.72e+00
20	62	41.9	380	1	TVFVMS	8.72e+00
21	62	41.9	602	1	TVRTRR	8.72e+00
22	62	41.9	647	1	S00644	8.72e+00
23	62	41.9	648	1	TVRTRR	8.72e+00

24	62	41.9	648	1	TVHUP6	8.72e+00
25	61	41.2	181	2	A55944	1.25e+01
26	61	41.2	202	1	HYVH42	1.25e+01
27	61	41.2	331	1	DEECG3	1.25e+01
28	61	41.2	331	1	DEURC3	1.25e+01
29	61	41.2	481	2	B32580	1.25e+01
30	61	41.2	1581	2	B71636	1.25e+01
31	60	40.5	142	2	A56552	1.78e+01
32	60	40.5	181	2	S38460	1.78e+01
33	60	40.5	184	2	JC2104	1.78e+01
34	60	40.5	210	1	JFVYX2	1.78e+01
35	60	40.5	221	2	R64858	1.78e+01
36	60	40.5	337	2	C64233	1.78e+01
37	60	40.5	342	2	A56552	1.78e+01
38	60	40.5	353	2	E64641	1.78e+01
39	60	40.5	471	2	S66829	1.78e+01
40	60	40.5	535	2	S76564	1.78e+01
41	60	40.5	664	2	PC4002	1.78e+01
42	60	40.5	692	2	H70362	1.78e+01
43	60	40.5	1066	2	F71100	1.78e+01
44	60	40.5	1177	2	A36984	1.78e+01
45	59	39.9	481	2	JC4342	2.54e+01

ALIGNMENTS

RESULT 1
 ENTRY B33135 #type complete
 TITLE gtfB protein precursor - Streptococcus mutans
 ORGANISM #formal_name Streptococcus mutans
 DATE 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 09-Sep-1997

ACCESSIONS B33135; A33128
 REFERENCE B33135
 #authors Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
 #journal J. Bacteriol. (1987) 169:4263-4270
 #title Sequence analysis of the gtfB gene from Streptococcus mutans.
 #cross-references NID:87308013
 #accession B33135
 #status Preliminary

##molecule_type DNA
 ##residues 1-1475 ##label SH1
 ##cross-references GB:M17361; NID:9153639; PID:9153640
 REFERENCE A33128
 #authors Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
 #submission submitted to the Protein Sequence Database, September 1990
 #accession A33128
 ##status Preliminary; not compared with conceptual translation
 ##molecule_type DNA
 ##residues 1-171,173-641,'N',643-1475 ##label SH2
 ##cross-references #superfamily cpl repeat homology

CLASSIFICATION #superfamily cpl repeat homology

FEATURE
 1096-1115 #domain cpl repeat homology #label CP1\
 1224-1243 #domain cpl repeat homology #label CP2\
 1289-1308 #domain cpl repeat homology #label CP3\
 1354-1373 #domain cpl repeat homology #label CP4\
 1419-1438 #domain cpl repeat homology #label CP5\
 #domain cpl repeat homology #label CP6\
 #length 1475 #molecular_weight 165811 #checksum 7497

SUMMARY
 Query Match 100.0%; Score 148; DB 2; Length 1475;
 Best local similarity 100.0%; Pred. No. 7.09e-17;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 481 ANDHLSILEAWSNDPYLHD 501
 1 ANDHLSILEAWSNDPYLHD 21
 RESULT 2
 ENTRY JT0345 #type complete
 TITLE dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans
 (strain GS-5)

```

ALTERNATE_NAMES  sucrose 6-glucosyltransferase
ORGANISM          #formal_name Streptococcus mutans
DATE              31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
                22-Nov-1996
ACCESSIONS       JT0345; C33135
REFERENCE         #authors Ueda, S.; Shiroza, T.; Kuramitsu, H.K.
                  #journal Gene (1988) 69:101-109
                  #title Sequence analysis of the gtfC gene from Streptococcus mutans
                  #accession GS-5.
                  #cross-references MUID:89137980
                  #accession JT0345
                  #molecule_type DNA
                  #residues 1-1375 ##label UED
                  ##experimental_source GS-5
REFERENCE         #authors Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
                  #journal J. Bacteriol. (1987) 169:4263-4270
                  #title Sequence analysis of the gtfB gene from Streptococcus mutans.
                  #cross-references MUID:87308013
                  #accession C33135
                  #status Preliminary
                  #molecule_type DNA
                  #residues 1-1349 ##label SHI
                  ##cross-references GB:M17361
GENETICS          #gene gtfC
FUNCTION          #description catalyzes the synthesis of both water-soluble and
                  #water-insoluble glucans from glucose.
CLASSIFICATION   #superfamily cpl repeat homology
KEYWORDS          #duplication: glycosyltransferase; hexosyltransferase;
                  #transferase
FEATURE          1-34
                  35-1375
                  #domain signal sequence #status predicted #label SIG\
                  #product glucosyltransferase #status predicted #label
                  #MAT\
1126-1135         #domain cpl repeat homology #label CP1\
1253-1272         #domain cpl repeat homology #label CP2\
1318-1337         #domain cpl repeat homology #label CP3
SUMMARY          #length 1375 #molecular-weight 153021 #checksum 7015
Query Match      91.9%; Score 136; DB 2; Length 1375;
Best Local Similarity 95.2%; Pred. No. 3,13e-14;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 507 ANDHLSTLEAMSDNDPYLHD 527
1 ANDHLSTLEAMSDNDPYLHD 21
RESULT 3
ENTRY 3
TITLE A38175 #type complete
ORGANISM #formal_name Streptococcus sobrinus
DATE 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change
09-Sep-1997
ACCESSIONS A38175
REFERENCE #authors Abo, H.; Mtsunura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato,
          K.; Kagawa, H.
          #journal J. Bacteriol. (1991) 173:989-996
          #title Peptide sequences for sucrose splitting and glucan binding
          #within Streptococcus sobrinus glucosyltransferase
          #water-insoluble glucan synthetase).
          #cross-references MUID:91123227
          #accession A38175
          #status Preliminary
          #molecule_type DNA
          #residues 1-1592 ##label ABO
          ##cross-references GB:D90213; NID:g217032; PID:d1014946; PID:g217033
CLASSIFICATION #superfamily cpl repeat homology
FEATURE

```

```

1093-1112         #domain cpl repeat homology #label CP1\
1222-1241         #domain cpl repeat homology #label CP2\
1287-1306         #domain cpl repeat homology #label CP3\
1330-1351         #domain cpl repeat homology #label CP4\
1352-1371         #domain cpl repeat homology #label CP5\
1402-1420         #domain cpl repeat homology #label CP6\
1465-1484         #domain cpl repeat homology #label CP7\
1513-1532         #domain cpl repeat homology #label CP8
SUMMARY          #length 1592 #molecular-weight 176167 #checksum 5940
Query Match      88.5%; Score 131; DB 2; Length 1592;
Best Local Similarity 85.7%; Pred. No. 3,81e-13;
Matches 18; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 477 ANNEVSIWEAMSDNDPYLHD 497
1 ANDHLSTLEAMSDNDPYLHD 21
RESULT 4
ENTRY 4
TITLE A41483 #type complete
ORGANISM #formal_name Streptococcus sobrinus
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
09-Sep-1997
ACCESSIONS A41483
REFERENCE #authors Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.
          #journal Infect. Immun. (1990) 58:2452-2458
          #title Analysis of the Streptococcus downei gtfS gene, which
          #specifies a glucosyltransferase that synthesizes soluble
          #glucans.
          #cross-references MUID:90316665
          #accession A41483
          #molecule_type DNA
          #residues 1-1365 ##label GIL
          ##cross-references GB:M30943; NID:g153652; PID:g153653
GENETICS          #gene gtfS
CLASSIFICATION   #superfamily cpl repeat homology
KEYWORDS          #glycosyltransferase; hexosyltransferase
SUMMARY          #length 1365 #molecular-weight 151590 #checksum 823
Query Match      68.9%; Score 102; DB 2; Length 1365;
Best Local Similarity 71.4%; Pred. No. 4,40e-07;
Matches 15; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
Db 467 AIDHLSTLEAMSGNDNDYKRD 487
1 ANDHLSTLEAMSDNDPYLHD 21
QY 1 ANDHLSTLEAMSDNDPYLHD 21
RESULT 5
ENTRY 5
TITLE A45866 #type complete
ORGANISM #formal_name Streptococcus mutans
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
07-Jul-1995
ACCESSIONS A45866
REFERENCE #authors Honda, O.; Kato, C.; Kuramitsu, H.K.
          #journal J. Gen. Microbiol. (1990) 136:2099-2105
          #title Nucleotide sequence of the Streptococcus mutans gtfD gene
          #encoding the glucosyltransferase-S enzyme.
          #accession A45866
          #status Preliminary
          #molecule_type DNA
          #residues 1-1431 ##label HON
          ##cross-references GB:M29296
CLASSIFICATION #superfamily cpl repeat homology
KEYWORDS          #glycosyltransferase; hexosyltransferase
FEATURE

```

181-201 #domain cpl repeat homology #label CP1
1127-1146 #domain cpl repeat homology #label CP2
1192-1211 #domain cpl repeat homology #label CP3
1257-1276 #domain cpl repeat homology #label CP4
1277-1297 #domain cpl repeat homology #label CP5
1321-1340 #domain cpl repeat homology #label CP6
1341-1361 #domain cpl repeat homology #label CP7
1385-1404 #domain cpl repeat homology #label CP7
SUMMARY #length 1431 #molecular-weight 159844 #checksum 4333

Query Match 68.2%: Score 101; DB 2; Length 1431;
Best Local Similarity 77.8%; Pred. No. 6.99e-07;
Matches 14; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

DB 495 AINHLSEAWMSNDPXY 512
1 :|||||
OY 1 ANDHLSLEAWMSNDPXY 18

RESULT 6
ENTRY S22737 #type complete
TITLE glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius
ORGANISM #formal_name Streptococcus salivarius
DATE 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change
09-Sep-1997
ACCESSIONS S22737; S28810; B44811; S22727
REFERENCE S22726
#authors Jacques, N.
#submission submitted to the EMBL Data Library, March 1992
#accession S22737
#molecule_type DNA
#residues 1-1599 #label JAC
#cross-references EMBL:Z11872; NID:g47530; PID:g47531
#experimental_source ATCC 25975

REFERENCE A44811
#authors Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.
#journal J. Gen. Microbiol. (1991) 137:2577-2593
#title Molecular characterization of a cluster of at least two glucosyltransferase genes in Streptococcus salivarius ATCC 25975.

#cross-references MVID:92148377
#accession S28810
#molecule_type DNA
#residues 1-51 #label GIF
#cross-references EMBL:Z11873

GENETICS
#gene gtfK
CLASSIFICATION #superfamily cpl repeat homology
KEYWORDS glucosyltransferase; hexosyltransferase
SUMMARY #length 1599 #molecular-weight 176480 #checksum 3135

Query Match 52.7%: Score 78; DB 2; Length 1599;
Best Local Similarity 55.6%; Pred. No. 1.70e-02;
Matches 10; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

DB 494 NISLEAWSHNDPYVNE 511
:|||||
OY 4 HSLSEAWMSNDPXYLH 21

RESULT 7
ENTRY A44811 #type complete
TITLE glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius
ORGANISM #formal_name Streptococcus salivarius
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
09-Sep-1997
ACCESSIONS A44811; S22726; S28809
REFERENCE A44811
#authors Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.
#journal J. Gen. Microbiol. (1991) 137:2577-2593
#title Molecular characterization of a cluster of at least two glucosyltransferase genes in Streptococcus salivarius ATCC 25975.

#cross-references MVID:92148377
#accession A44811
#molecule_type DNA
#residues 1-1518 #label GIF
#cross-references EMBL:Z11873; NID:g47526; PID:g47527
#note Sequence extracted from NCBI Backbone (NCBIN:81050, NCBI:81052)

GENETICS
#gene gtfJ
CLASSIFICATION #superfamily cpl repeat homology
KEYWORDS glucosyltransferase; hexosyltransferase
SUMMARY #length 1518 #molecular-weight 167730 #checksum 1334

Query Match 50.7%: Score 75; DB 2; Length 1518;
Best Local Similarity 66.7%; Pred. No. 5.82e-02;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

DB 504 HISVLEAWSLDNHXY 518
:|||||
OY 4 HSLSEAWMSNDPXY 18

RESULT 8
ENTRY B48445 #type complete
TITLE glyceraledehyde-3-phosphate dehydrogenase (EC 1.2.1.12) -
ORGANISM Leishmania mexicana
#formal_name Leishmania mexicana
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
08-Sep-1997
ACCESSIONS B48445; S25142
REFERENCE B48445
#authors Hanneert, V.; Blaauw, M.; Kohl, L.; Allert, S.; Oppendoes, F.R.; Michels, P.A.M.;
#journal Mol. Biochem. Parasitol. (1992) 55:115-126
#title Molecular analysis of the cytosolic and glycosomal glyceraledehyde-3-phosphate dehydrogenase in Leishmania mexicana.

#cross-references MVID:93063042
#accession B48445
#status preliminary
#molecule_type DNA
#residues 1-331 #label HAN
#cross-references EMBL:X65220; NID:g9552; PID:g9553

CLASSIFICATION #superfamily glyceraledehyde-3-phosphate dehydrogenase
KEYWORDS oxidoreductase
SUMMARY #length 331 #molecular-weight 35643 #checksum 147

Query Match 48.6%: Score 72; DB 2; Length 331;
Best Local Similarity 52.6%; Pred. No. 1.94e-01;
Matches 10; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

DB 301 NDHFVRLVSNYDNGYSH 319
:|||||
OY 2 NDHLSLEAWMSNDPXYLH 20

RESULT 9
ENTRY QOBE40 #type complete
TITLE BGLF2 protein - human herpesvirus 4 (strain B95-8)
ORGANISM #formal_name human herpesvirus 4, Epstein-Barr virus
DATE 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change
22-Jan-1999
ACCESSIONS C43044; J01381; A03784; A03794; S33036
REFERENCE A93065
#authors Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
#journal Mol. Biol. Med. (1983) 1:21-45
#title Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus.
#cross-references MVID:85035713
#accession C43044
#molecule_type DNA

##residues 1-336 ##label BAN
##cross-references EMBL:V01555; NID:959074; PID:g1334895
REFERENCE A03794
#authors Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.;
Farrell, P.J.; Gibson, T.J.; Hatfull, G.; Hudson, G.S.;
Satchwell, S.C.; Seglin, C.; Tiffinell, P.S.; Barrell, B.G.
#journal Nature (1984) 310:207-211
#title DNA sequence and expression of the B95-8 Epstein-Barr virus
genome.
#cross-references MUID:84270667
#contents annotation: protein coding region
#note neither amino acid nor nucleotide sequence is given
REFERENCE J01381
#authors Chen, M.R.; Hsu, T.Y.; Lin, S.W.; Chen, J.Y.; Yang, C.S.
#journal J. Gen. Virol. (1991) 72:3047-3055
#title Cloning and characterization of cDNA clones corresponding to
transcripts from the BamHI G region of the Epstein-Barr
virus genome and expression of BGLF2.
#cross-references MUID:92113548
accession J01381
#molecule_type mRNA
##residues 1-336 ##label CHE
##cross-references GB:S77132; NID:9243314; PID:9243315
CLASSIFICATION #superfamily Epstein-Barr virus BGLF2 protein
SUMMARY #length 336 #molecular-weight 36888 #checksum 8731
Query Match 48.6%; Score 72; DB 1; Length 336;
Best Local Similarity 41.2%; Pred. No. 1.94e-01;
Matches 7; Conservative 7; Mismatches 3; Indels 0; Gaps 0;
DB 216 AGAIVNLRGWTEDDSP 232
QY 1 ANDHLEAMSDNDTP 17
RESULT 10
ENTRY E71557 #type complete
TITLE hypothetical protein CT085 - Chlamydia trachomatis (serotype
D, strain UW3/Cx)
ORGANISM #formal_name Chlamydia trachomatis
DATE 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change
21-Nov-1998
ACCESSIONS E71557
REFERENCE A71570
#authors Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe,
R.; Aravind, L.; Mitchell, W.P.; Olinger, L.; Tatusov,
R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.
#journal Science (1998) 282:754-759
#title Genome sequence of an obligate intracellular pathogen of
humans: Chlamydia trachomatis.
#cross-references MUID:9900809
accession E71557
#status preliminary
##molecule_type DNA
##residues 1-579 ##label ARN
##cross-references GB:AE001283; GB:AE001273; NID:93328480; PID:93328481
#experimental_source serotype D, strain UW-3/Cx
GENETICS CT085
#gene #length 579 #molecular-weight 65354 #checksum 8323
SUMMARY
Query Match 48.6%; Score 72; DB 2; Length 579;
Best Local Similarity 31.3%; Pred. No. 1.94e-01;
Matches 5; Conservative 8; Mismatches 3; Indels 0; Gaps 0;
DB 132 QLPMLTSMPEDCGFL 147
QY 4 HLESLAMSDNDTPYL 19
RESULT 11
ENTRY H70538 #type complete
TITLE Probable ppdk protein - Mycobacterium tuberculosis (strain

ORGANISM H37RV)
#formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
17-Jul-1998
ACCESSIONS H70538
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eigemeier, K.; Gas, S.; Barry
III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skellton, S.; Squares, S.; Squares, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from
the complete genome sequence.
#cross-references MUID:98295987
accession H70538
#status preliminary: nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-490 ##label COL
##cross-references GB:595585; GB:AL123456; NID:93261787; PID:e317147;
PID:g2117221
#experimental_source strain H37RV
GENETICS
#gene #length 490 #molecular-weight 52004 #checksum 20
SUMMARY
Query Match 45.3%; Score 67; DB 2; Length 490;
Best Local Similarity 57.1%; Pred. No. 1.36e+00;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
DB 467 LSAMSEDPPELRE 480
QY 8 LEAMSDNDTPYLHD 21
RESULT 12
ENTRY JC4086 #type complete
TITLE fatty-acid synthase (Ec 2.3.1.85) alpha chain - yeast
(Candida albicans)
ORGANISM #formal_name Candida albicans
DATE 12-Jul-1995 #sequence_revision 03-Aug-1995 #text_change
05-Dec-1998
ACCESSIONS JC4086
REFERENCE JC4086
#authors Southard, S.B.; Cihlar, R.L.
#journal Gene (1995) 156:133-138
#title Analysis and expression of the Candida albicans FAS2 gene.
#cross-references MUID:95255657
accession JC4086
#molecule_type DNA
##residues 1-1885 ##label SOU
##cross-references GB:L29063; NID:g456442; PID:g456443
GENETICS
#gene #length 1885
#map_position 3
CLASSIFICATION #superfamily yeast fatty-acid synthase
KEYWORDS acyltransferase; fatty acid biosynthesis; phosphopantetheine;
phosphoprotein
FEATURE 1299-1304
181 #region cerulenin binding #status predicted
#binding_site phosphopantetheine (Ser) (covalent)
#status predicted
SUMMARY #length 1885 #molecular-weight 207588 #checksum 7273
Query Match 44.6%; Score 66; DB 2; Length 1885;
Best Local Similarity 25.0%; Pred. No. 1.98e+00;
Matches 5; Conservative 8; Mismatches 7; Indels 0; Gaps 0;


```

Db      1452  NKRLEQIKSWEETELSYLQE 1471
          | : | : : : : : : : : :
QY      2    NDHLSILEAWSDNDTPYLHD 21

```

QY 2 NDHLSILEAWSNDNDTPYLHD 21

RESULT	13
ENTRY	S56229
TITLE	#type complete
ORGANISM	pcha protein - Pseudomonas aeruginosa
DATE	#formal_name Pseudomonas aeruginosa
	13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change

```

##status      preliminary
##molecule_type  DNA
##residues    1-476 ##label SE2
##cross-references EMBL:X82644: NID:g1628425: PID:g929780
##note        this protein is involved in salicylate biosynthesis
CLASSIFICATION #superfamily 1 isochorismate synthase
SUMMARY        #length 476 #molecular_weight 52071 #checksum 7909

```

ENTRY	14
RESULT	
TITLE	A30065 #type complete
CONTAINS	trigramin precursor - Indian green tree viper hemorrhagic proteinase (EC 3.4.24.-); platelet aggregation inhibitor (disintegrin) trigramin
ORGANISM	#formal_name Trimeresurus gramineus #common_name Indian green tree viper
DATE	16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 05-Sep-1997
ACCESSIONS	S12589; A30065; A29784
REFERENCE	S12589
^authors	Neeper, M.P.; Jacobson, M.A.
^journal	Nucleic Acids Res. (1990) 18:425
^title	Sequence of a cDNA encoding the platelet aggregation inhibitor trigramin.
#cross-references	MUDD:90332429
#accession	S12589

```

REFERENCE
#authors      A30065
#journal      Huang, T.F.; Holt, J.C.; Kirby, E.P.; Niewiarowski, S.
#title        Biochemistry (1989) 28:661-666
              Triptamin: primary structure and its inhibition of von
              Willebrand factor binding to glycoprotein IIb/IIIa complex
              on human platelets.
#cross-references MIMD:89228063
#accession    A30065
#molecule-type protein
#residues     408-479 ##label HVA
REFERENCE
#authors      A29784
#journal      Huang, T.F.; Holt, J.C.; Lukaszewicz, H.; Niewiarowski, S.
              J. Biol. Chem. (1987) 262:16157-16163

```

```
#title
    Trigramin. A low molecular weight peptide inhibiting
    fibrinogen interaction with platelet receptors expressed on
    glycoprotein IIb-IIIa complex.
#cross-references M0ID:88058981
#accession
    A29784
#description
    trnsfected
```

```
##molecule_type protein
##residues 408-419 ##label HU2
CLASSIFICATION #superfamily trigramin precursor; disintegrin homology
KEYWORDS anticoagulant; glycoprotein; hydrolase; metalloproteinase
          venom; zinc; zymogen
```

```

1-18      #domain signal sequence #status predicted #label SIG
398-476   #domain distintegrin homology #label DIS\
408-479   #product trigramin #status experimental #label MAT\
458-460   #region cell attachment (R-G-X) motif\
279       #binding site carbohydrate (Asn) (covalent) #status
          predicted\
333,337,343 #binding_site zinc (His) #status predicted\
334         #active_site glu #status predicted\
SUMMARY    #length 480 #molecular_weight 53454 #checksum 1021

```

ENTRY TITLE	RESULT	15
<p> #type complete E71447 probable PSTII D1 protein processing enzyme - Arabidopsis thaliana #formal_name Arabidopsis thaliana #common_name mouse-ear cress columbia 03-Aug-1998 #sequence-revision 03-Aug-1998 #text-change 03-Dec-1998 w17 </p>		

```

GENETICS
  #map_position 4COP9-4G3845
  #molecule_type DNA
  ##residues 1-515 ##label BEV
  ##crosso-references GB:827344; NID:g2245126; PID:e327081; PID:g2245133
SUMMARY
  #length 515 #molecular-weight 55762 #checksum 303
Query Match 43.2%; Score 64; DB 2; Length 515;
Best Local Similarity 36.8%; Pred. No. 4.19e+00;

```

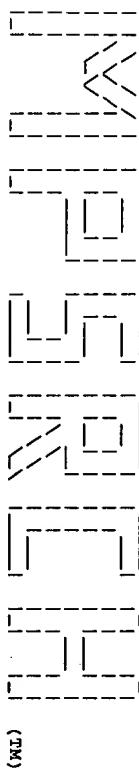
Thu Jan 13 09:13:33 2000

US-09-290-049-1.rpr

Page 6

```
Matches 7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;
Db 128 TEENLEFLEARNTRIDRAYI 146
:::|:||||| |:|:
QY 1 ANDHLSLEAWSNDNDPYL 19
.
```

Search completed: Tue Jan 11 15:37:09 2000
Job time : 14 secs.



Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Jan 11 15:46:28 2000; Maspar time 3.84 Seconds
121.776 Million cell updates/sec
ar output not generated.

Title: >US-09-290-049-3
Description: (1-22) from US09290049.pep
Perfect Score: 155
Sequence: 1 TGARTINGQLYFRANGVQVKG 22

Scoring table: PAM 150
Gap 15

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:

a-geneseg35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 20.775; Variance 64.431; scale 0.322

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	114	73.5	22	25	W34159	GTF antigenic peptide	4.41e-05
2	114	73.5	22	8	R43696	GTF subsequence.	4.41e-05
3	114	73.5	24	25	W34164	GTF antigenic peptide	4.41e-05
4	112	72.3	15	2	R32925	Glucosyltransferase	7.62e-05
5	88	56.8	15	17	R91047	Alpha-D-glucosyltrans	4.64e-02
6	67	43.2	8	11	R95014	C. difficile toxin A	9.21e+00
7	67	43.2	8	12	R95017	C. difficile toxin A	9.21e+00
8	67	43.2	8	17	R95016	C. difficile toxin A	9.21e+00
9	67	43.2	8	35	W68387	Clostridium difficile	9.21e+00
10	63	40.6	16	19	R78621	Chicken galNAc-4-epi	3.04e+01
11	62	40.0	16	19	W06416	Phosphotriesterase-re	3.04e+01
12	61	39.4	34	3	W38549	S. pneumoniae peptide	3.85e+01
13	61	39.4	34	39	W83359	Streptococcus pneumon	3.85e+01
14	60	38.7	19	16	W03642	Human cannabinoid GPR	4.87e+01
15	60	38.7	19	16	R14196	Human cannabinoid rec	4.87e+01
16	60	38.7	19	16	R14196	Rat cannabinoid recep	4.87e+01

17	59	38.1	91	17	R95015	C. difficile toxin A	6.14e+01
18	59	38.1	23	13	R67918	(1-3)-beta-D-glucan s	6.14e+01
19	59	38.1	45	10	R51083	fabg gene encoding bl	6.14e+01
20	59	38.1	45	20	W05208	Synechococcus biofin	6.14e+01
21	59	38.1	45	35	W70403	Synechococcus biofin	6.14e+01
22	59	38.1	67	13	R67915	(1-3)-beta-D-glucan s	6.14e+01
23	59	38.1	99	19	W03421	Mouse developmental k	6.14e+01
24	59	38.1	98	15	R85092	EPH-like receptor pro	6.14e+01
25	58	37.4	1301	6	R28304	Acid alpha-amylose en	9.74e+01
26	57	36.8	488	5	R24862	Sequence of pre-pro s	9.74e+01
27	57	36.8	492	5	R24863	Sequence of pre-pro s	9.74e+01
28	57	36.8	501	15	R89278	Ca-cellulase.	9.74e+01
29	57	36.8	501	3	R14372	Tomato endo-1,4-B-glu	9.74e+01
30	57	36.8	572	8	R40843	Bilirubin oxidase.	9.74e+01
31	56	36.1	15	9	R45658	Human progesterone re	1.22e+02
32	56	36.1	336	6	R34293	C-terminally truncated	1.22e+02
33	56	36.1	336	6	R34293	C-terminally truncated	1.22e+02
34	56	36.1	432	6	R34294	HPL(-) mutant with GP	1.22e+02
35	56	36.1	448	9	R49873	Bacillus sphaericus 5	1.22e+02
36	56	36.1	450	6	R30738	Human pancreatic lipa	1.22e+02
37	56	36.1	466	22	W20586	H. pylori cytoplasmic	1.22e+02
38	55	35.5	265	28	W46518	Amino acid sequence o	1.53e+02
39	55	35.5	449	12	R66742	Biotin-carboxylase.	1.53e+02
40	55	35.5	485	3	R13513	P.denitrificans COB Q	1.53e+02
41	55	35.5	503	3	P70219	Sequence of serrapept	1.53e+02
42	55	35.5	1139	37	W76425	Human JAK2 protein se	1.53e+02
43	55	35.5	1657	24	W18822	Human ICGAP1.	1.92e+02
44	54	34.8	667	28	W44701	Chimeric Cre-LDB-GR(I	1.92e+02
45	54	34.8	236	17	R95011	C. difficile toxin B.	1.92e+02

ALIGNMENTS

RESULT 1
ID W34159 standard; peptide: 22 AA.
AC W34159; 18-FEB-1998 (first entry)
DT 18-FEB-1998 (first entry)
DE GTF antigenic peptide #4.
KW Glucosyltransferase; GTF; enzyme; immunogen; catalytic domain; vaccine;
KW surface domain; glucan-binding domain; mutans streptococcal strain;
KW immune response; glucan-binding activity; dental caries prevention.
OS Streptococcus mutans.
PN US5686075-A.
PD 11-NOV-1997.
PF 01-MAY-1992; 877295.
PR 30-APR-1993; US-057162.
PR 01-MAY-1992; US-877295.
PA (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.
PI Smith DJ, Taubman MA;
DR WPI: 97-558089/51.
PT Immunogenic compositions containing streptococcal
PT glucosyl:transferase peptide(s) - used for provoking immune response
PT to streptococcal glucosyl:transferase for preventing dental caries
PT Claim 2; Column 13; 11pp: English.
CC W34156-W34160 represent immunogenic fragments of the Streptococcus mutans
CC glucosyltransferase (GTF) enzyme. W34157 and W34158 are from the
CC catalytic domain of GTF, while W34160 is from the GTF surface domain.
CC W34156 and W34159 are from the glucan-binding domain of GTF. These
CC sequences can all be used in the immunogenic composition of the
CC invention. The composition of the invention can alternatively comprise
CC one or more of these sequences linked to a lysine core matrix (see
CC W34161-W34165). A composition comprising one of these sequences can be
CC administered to a mammal to raise an immune response, in a method for
CC interfering with the enzymatic activity of streptococcal
CC glucosyltransferase in a mammal. The immune response results in reduction
CC of the colonisation or accumulation of mutans streptococcal strains in
CC the mammal. Compositions containing W34156 specifically interfere with
CC the glucan-binding activity of the streptococcal glucosyltransferase.
CC The peptides can also be used in vaccines for preventing dental caries
CC in mammals. 22 AA:
Sequence
Query Match 73.5%; Score 114; DB 25; Length 22;

QY 1 TGARTINGQLXFRANGVOYKG 22

RESULT 5
ID R91047 standard; Protein: 1577 AA.

AC R91047:
DE 22-MAY-1996 (first entry)
KW Alpha-D-glucosyltransferase.
KW sucrose; transgenic plant; cloning; Escherichia coli;
KW phase lambda-cl3; vector; plasmid pGS6501; plasmid pGS6502;
KW gene transfer; crop improvement; storage carbohydrate; pasture;
KW feedstuff; senescence; dextran; binder; food; pharmaceutical.
OS Streptococcus salivarius strain ATCC 25975.
PN WO9606173-A1.
PD 29-FEB-1996.
PR 24-AUG-1995: AU0527.
PR 24-AUG-1994: AU-007643.
PA (GIFP/) GIFFARD P.M.
PA (JACO/) JACOUES N.A.
PA (SIMP/) SIMPSON C.L.
F Giffard PM, Jacques NA, Simpson CL;
D WPI: 96-151376/15.
DR N-PSDB: 113139.
PT Plants contg. new bacterial DNA encoding glucosyl transferase
PT activity - retain higher levels of stored carbohydrate(s) in a form
PT readily digestible by ruminants
PS Claim 4; Page 16-20; 31pp; English.
CC The sequence represents an alpha-D-glucosyltransferase from
CC Streptococcus salivarius. The enzyme is primer-independent, and
CC produces soluble glucan from sucrose. A gene encoding the enzyme
CC may be cloned and expressed in Escherichia coli using a subclone
CC of phage lambda-cl3, e.g. plasmid pGS6501 or plasmid pGS6502. The
CC DNA may also be expressed in a transgenic plant, to improve the
CC level of stored carbohydrate in a pasture plant which normally
CC contains low levels, or to prevent degradation of stored carbohydrate
CC during plant senescence. Dextran may be isolated from the plant, for
CC use as a food binder or pharmaceutical additive. Primer independence
CC ensures that the enzyme will be functional in plants. The glucan is
CC poorly degraded in plants but easily degraded by bacteria in the rumen
CC of grazing livestock.
CC Sequence 1577 AA;
SQ

Query Match 56.8%; Score 88; DB 16; Length 1577;
Best Local Similarity 59.1%; Pred. No. 4.64e-02;

Matches 13; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

DB 1533 tggvtnqkqlyfdsggrqyk 1554

1 TGARTINGQLXFRANGVOYKG 22

RESULT 6

ID R95014 standard; Protein: 811 AA.

AC R95014:
DE 08-JUL-1996 (first entry)
C. difficile toxin A (aal870-2680).
KW Toxin A; cytotoxin; enterotoxin; fusion protein; antitoxin;
KW diarrhoea; diagnosis; therapy.
OS Clostridium difficile VPI strain 10463 (ATCC 10463).
PN WO9612802-A1.
PD 02-MAY-1996.
PF 23-OCT-1995: U13737.
PR 24-OCT-1994: US-329154.
PR 16-MAR-1995: US-405496.
PR 14-APR-1995: US-422711.
PR 07-JUN-1995: US-480604.
PA (OPHI-) OPHIDIAN PHARM INC.
PA Flicra JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;
PI Williams JA;
DR WPI: 96-230603/23.
PT Fusion proteins comprising non-toxin protein and part of toxin
PT useful to form anti-toxins against Clostridium botulinum type A, and

PT C. difficile type toxins, and to treat C. difficile intoxication,
PT partic. diarrhoea

PS Claim 53; Page 310-312; 434pp; English.
CC A recombinant protein (R95014) comprises amino acids 1870-2680 of
CC Clostridium difficile VPI strain 10463 toxin A (see also R95016),
CC a cytotoxin associated with diarrhoeic disease. It was obt. by
CC PCR amplification of genomic DNA, and was expressed as a soluble
CC fusion protein, with maltose binding protein as fusion partner, in
CC transformed E. coli host cells. The soluble fusion protein can be
CC used to raise avian antibodies useful as antitoxins or diagnostics.
SQ Sequence 811 AA;

Query Match 43.2%; Score 67; DB 17; Length 811;
Best Local Similarity 52.2%; Pred. No. 9.21e-00;

Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

DB 587 tglrtldgkkyfntntravcg 609

1 TGARTINGQLXFRAN-GVOYKG 22

RESULT 7

ID R95017 standard; Protein: 812 AA.

AC R95017:
DE 08-JUL-1996 (first entry)
C. difficile toxin A (aal873-2684).
KW Toxin A; cytotoxin; enterotoxin; fusion protein; antitoxin; vaccine;
KW immunogen; diarrhoea; diagnosis; therapy.
OS Clostridium difficile VPI strain 10463 (ATCC 10463).
PN WO9612802-A1.
PD 02-MAY-1996.
PF 23-OCT-1995: U13737.
PR 24-OCT-1994: US-329154.
PR 16-MAR-1995: US-405496.
PR 14-APR-1995: US-422711.
PR 07-JUN-1995: US-480604.
PA (OPHI-) OPHIDIAN PHARM INC.
PA Flicra JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;
PI Williams JA;
DR WPI: 96-230603/23.
PT Fusion proteins comprising non-toxin protein and part of toxin
PT useful to form anti-toxins against Clostridium botulinum type A, and
PT C. difficile type toxins, and to treat C. difficile intoxication,
PT partic. diarrhoea
PS Claim 79; Page 354-356; 434pp; English.
CC A protein (R95017) comprising amino acids 1873-2684 of Clostridium
CC difficile VPI strain 10463 toxin A (see also R95016) was obt. by
CC PCR amplification of genomic DNA. The PCR product can be expressed
CC in transformed E. coli host cells as a soluble fusion protein, with
CC maltose binding protein or a polystyrene affinity tag as
CC fusion partner. The resulting soluble fusion proteins, which are
CC substantially endotoxin-free, can be used in anti-clostridial
CC vaccines or to raise avian antibodies useful in clostridial antitoxin
CC therapy for humans and animals.
SQ Sequence 812 AA;

Query Match 43.2%; Score 67; DB 17; Length 812;
Best Local Similarity 52.2%; Pred. No. 9.21e-00;

Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

DB 588 tglrtldgkkyfntntravcg 610

1 TGARTINGQLXFRAN-GVOYKG 22

RESULT 8

ID R95016 standard; Protein: 2710 AA.

AC R95016:
DE 08-JUL-1996 (first entry)
C. difficile toxin A.
KW Toxin A; cytotoxin; enterotoxin; fusion protein; antitoxin;
KW vaccine; diarrhoea; diagnosis; therapy.
OS Clostridium difficile VPI strain 10463 (ATCC 10463).

PN W09612802-A1.
 PD 02-MAY-1996.
 PF 23-OCT-1995; U13737.
 PR 24-OCT-1994; US-329154.
 PR 16-MAR-1995; US-405496.
 PR 14-APR-1995; US-422711.
 PR 07-JUN-1995; US-480604.
 PA (OPH1-) OPHIDIAN PHARM INC.
 PI Filica JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;
 PI Williams JA;
 DR WPI: 96-230603/23.
 DR N-PSDB: T29248.
 PT Fusion proteins comprising non-toxin protein and part of toxin
 PT useful to form anti-toxins against Clostridium botulinum type A, and
 PT C. difficile type toxins, are used to treat C. difficile intoxication,
 PT paritc. diarrhoea
 PS Claim 63: Page 290-302; 434pp; English.
 CC Clostridium difficile VPI strain 10463 toxin A (R95016), the
 CC product of the toxin A gene (T29248), is a potent cytotoxin that
 CC plays a direct role in damaging gastrointestinal tissues and is
 CC associated with diarrhoeic disease. It can be obt. by expression in
 CC transformed E. coli hosts of portions of DNA that together cover the
 CC entire toxin A gene. Toxin A, and portions of it (see also
 CC R95014-15 and R95017), pref. expressed as fusions to polystyridine
 CC affinity tags or maltose binding protein, are used to raise avian
 CC antibodies useful as antitoxins or diagnostics, and in vaccine prodn.
 CC Sequence 2710 AA;

Query Match 43.28; Score 67; DB 17; Length 2710;
 Best Local Similarity 52.28; Pred. No. 9.21e+00;

Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

Db 2460 tglrtldgkkyfntavtg 2482
 |||||:|||||
 OY 1 TGARTINGOLLYFRAN-GVOYKG 22

RESULT 9
 ID W68387 standard; Protein; 2710 AA.
 AC W68387;
 DT 07-DEC-1998 (first entry)
 DE Clostridium difficile toxin A.
 KM Antitoxin; vaccine; cytotoxin; toxin A; intoxication; immunogen;
 KM pseudomembranous enterocolitis.
 OS Clostridium difficile.
 PN W09808540-A1.
 PD 05-MAR-1998.
 PR 28-AUG-1997; U15394.
 PA (OPH1-) OPHIDIAN PHARM INC.
 PI Thalley BS, Williams JA;
 DR WPI: 98-230234/20.
 DR N-PSDB: V30560.
 PT Host cell containing recombinant expression vector encoding
 PT Clostridium botulinum type B or E toxin - useful to treat humans
 PT and other animals at risk of intoxication with clostridial toxin
 PS Example 15: Page 220-230; 428pp; English.
 CC This is the amino acid sequence of Clostridium difficile toxin A,
 CC deduced from the coding region (see V30560) of the toxin A gene.
 CC Toxin A is a potent cytotoxin that plays a direct role in damaging
 CC gastrointestinal tissues. Severe cases of C. difficile
 CC intoxication result in pseudomembranous colitis. This would be
 CC prevented by neutralising the effects of toxin A in the
 CC gastrointestinal tract. Examples are provided of the production
 CC of recombinant C. difficile toxin A in host cells and of the in
 CC vivo neutralisation of toxin A by antibodies against recombinant
 CC toxin A polypeptides. The invention specifically relates to
 CC recombinant proteins derived from Clostridium botulinum toxins
 CC (see W68389-400) and their use as immunogens for the production of
 CC vaccines and antitoxins.
 CC Sequence 2710 AA;

Query Match 43.28; Score 67; DB 35; Length 2710;

Best Local Similarity 52.28; Pred. No. 9.21e+00;
 Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;
 Db 2460 tglrtldgkkyfntavtg 2482
 |||||:|||||
 OY 1 TGARTINGOLLYFRAN-GVOYKG 22

RESULT 10
 ID R78621 standard; Protein; 404 AA.
 AC R78621;
 DT 11-JUN-1996 (first entry)
 DE Chicken GalNac-alpha-2,6-sialyltransferase P-B3.
 KM GalNac-alpha-2,6-sialyltransferase P-B3; anti-inflammatory;
 KM viral infection; tumour; migration; inhibitor; sugar chain;
 KM reagent; chicken.
 OS Gallus gallus.
 PN W09518217-A1.
 PD 06-JUL-1995.
 PF 22-DEC-1994; J02182.
 PR 24-DEC-1993; JP-348260.
 PR 28-MAR-1994; JP-057369.
 PR 28-APR-1994; JP-091507.
 PA (RIKA) INST PHYSICAL & CHEM RES.
 PI Hamamoto T, Kojima N, Kurosawa N, Lee Y, Nakaoka T;
 PI Tsuji S;
 DR WPI: 95-246383/32.
 DR N-PSDB: 095325.
 PT New GalNac-alpha-2,6-sialyltransferases P-B1 and P-B3 - for e.g.
 PT treatment of genetic disorders involving missing sugars
 PS Claim 4: Pages 52-55; 70pp; Japanese.
 CC 095325 encodes R78621 chicken GalNac-alpha-2,6-sialyltransferase
 CC P-B3. P-B3 can be used as a reagent for introducing human type
 CC sugar chains onto proteins, or for the treatment of genetic
 CC disorders involving missing sugar chains. It may also be used as
 CC a tumour migration inhibitor, viral infection preventative and
 CC as an anti-inflammatory.
 CC Sequence 404 AA;

Query Match 40.68; Score 63; DB 16; Length 404;
 Best Local Similarity 45.08; Pred. No. 2.40e+01;

Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Db 196 grahdahdvfringaltg 215
 :||::|||
 OY 3 ARTINGOLLYFRAN-GVOYKG 22

RESULT 11
 ID W06416 standard; Protein; 349 AA.
 AC W06416;
 DT 28-JAN-1997 (first entry)
 DE Phosphotriesterase-related protein.
 KM Phosphotriesterase-related protein; mouse; PPP; recessive cpk; murine;
 KM polycystic kidney disease; PKD; acute renal injury; autosomal dominant;
 KM autosomal recessive; fluid-filled cyst; nephron; collecting duct; kidney;
 KM renal parenchyma; renal failure; mammal; diagnostic marker; therapy;
 KM acute organophosphate toxicity; phosphotriesterase inhibitor;
 KM anti-cancer drug enhancer; anti-HIV drug enhancer.
 OS Mus musculus.
 PN U5552313-A.
 PD 03-SEP-1996.
 PF 21-NOV-1994; 343027.
 PR 21-NOV-1994; US-343027.
 PA (UNITV) UNITV KANSAS.
 PI Calvet JP, Hou X, Magenheimer BS, Maser RL;
 DR WPI: 96-412067/41.
 DR N-PSDB: T43210.
 PT Mouse phosphotriesterase-related protein DNA - used to develop
 PT probe. for diagnosis of polycystic kidney disease or acute renal
 PT failure
 PS Claim 1: Column 25-28; 23pp; English.
 CC This sequence represents the mouse phosphotriesterase related protein

(mPRP). The mPRP sequence has abnormal underexpression in the recessive cpx murine model of a polycystic kidney disease (PKD), and has decreased expression following acute renal injury. PKD is a common disease (affecting 1 in 500-1000 individuals) and can be inherited as an autosomal dominant or as an autosomal recessive. PKD is characterised by the development of innumerable, large, fluid-filled cysts from the nephrons and collecting ducts of affected kidneys. It is thought that enlargement of the cysts interferes with functioning of the normal renal parenchyma, which eventually leads to renal failure. The mPRP, or other similar mammalian PRPs can be used as diagnostic markers for PKD and acute renal failure. The PRPs can also be used in therapeutic and protective treatments, such as for acute organophosphate toxicity. They can also be used to develop products to enhance the effectiveness of other types of therapy, e.g. phosphodiesterase inhibitors to enhance the effectiveness of certain anti-cancer, or anti-HIV drugs.

Sequence 349 AA;

Query Match 40.08; Score 62; DB 19; Length 349;
Best Local Similarity 53.88; Pred. No. 3.04e+01;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

79 alreellykayg 91
: | |||||
5 TINGOLLYFRANG 17

RESULT 12

ID W38549 standard; Protein: 207 AA.

AC W38549;

DE 06-NOV-1998 (first entry)

KW Streptococcus peptidase chain release factor 3 (RF-3).

KW Immunological response; inoculation; antibody production; inhibitor;

KW T cell immune response; antimicrobial compound; bacterial adhesion;

KW extracellular matrix protein; protein-mediated cell invasion; wound;

KW pathogenesis.

OS Streptococcus pneumoniae.

PN MO9743303-A1.

PD 20-NOV-1997.

PF 14-MAY-1997; U07950.

PR 14-MAY-1996; US-017670.

PA (SMK) SMITHKLINE BEECHAM CORP.

PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO,

PI Stodola RK;

DR WPI: 98-008793/01.

DR N-PSDB: T98605.

PT Novel Streptococcus pneumoniae proteins and related DNA - useful for

PT diagnosing anti-microbial agents for treatment of bacterial

PT infections

PT Claim 12; Page 327; 483pp; English.

CC This sequence represents a Streptococcus pneumoniae protein that, based

CC on homology with Bacteroides nodosus protein, is a peptidase chain release

CC factor 3 (RF-3), and is encoded by a DNA sequence of the invention.

CC The DNA sequences were isolated from Streptococcus pneumoniae strain

CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the

CC invention can be used to identify compounds which interact with and

CC inhibit or activate the activity of the proteins. Antagonists can be

CC used to treat diseases caused by S. pneumoniae proteins, through genetic

CC immunisation. They can also be used to induce an immunological response

CC in a mammal by inoculation with the S. pneumoniae proteins or delivery

CC and/or T cell immune responses to protect the animal from disease. The

CC proteins can also be used to identify antimicrobial compounds which are

CC capable of inhibiting their bioactivity. In particular the proteins of

CC the invention can be used to prevent adhesion of bacteria to mammalian

CC extracellular matrix proteins on in-dwelling devices or in wounds, to

CC block protein-mediated mammalian cell invasion, and to block the normal

CC progression of pathogenesis in infections initiated other than by the

CC implantation of in-dwelling devices or other surgical techniques.

Sequence 207 AA;

Query Match 39.48; Score 61; DB 34; Length 207;

Best Local Similarity 61.58; Pred. No. 3.85e+01;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db 21 agtltteqlylf 33
: | || |||||
1 TGARTINGOLLYF 13

RESULT 13

ID W83359 standard; Protein: 514 AA.

AC W83359;

DE 17-FEB-1999 (first entry)

KW Streptococcus pneumoniae prfC protein sequence.

KW Streptococcus pneumoniae; prfC; antibacterial; meningitis; pneumonia;

KW peptide releasing factor; otitis media; conjunctivitis; bacteremia;

KW sinusitis; pleural empyema; endocarditis; immunisation; infection;

KW gene therapy.

OS Streptococcus pneumoniae.

PN EP-881292-A2.

PD 02-DEC-1998.

PF 26-MAY-1998; 304157.

PR 29-MAY-1997; US-865311.

PA (SMK) SMITHKLINE BEECHAM CORP.

PI Pearson SC;

PI WPI: 99-001398/01.

DR N-PSDB: V72588.

PT New Streptococcus pneumoniae peptide releasing factor polypeptide

PT and polynucleotide - useful as diagnostic reagents and for

PT prevention and treatment of diseases caused by bacterial infections,

PT including meningitis and pneumonia

PT Claim 12; Page 24-26; 27pp; English.

CC The present sequence is Streptococcus pneumoniae peptide releasing

CC factor (prfC). prfC polypeptides and polynucleotides are useful for

CC diagnosing susceptibility to diseases by detecting mutations or

CC polymorphisms of the prfC gene. PCR using prfC probes is useful for

CC diagnosing diseases caused by organisms comprising the prfC gene by

CC detection at the nucleic acid level, and analysing the presence or

CC amount of prfC polypeptide in cell or tissue samples. This method is

CC useful for diagnosing the stage of infection and the type of pathogen.

CC prfC polypeptides and polynucleotides can be used to screen for

CC antagonists and agonists (especially bacteriostatic and bacteriocidal

CC compounds), which can be used in treatment to enhance or block prfC

CC activity. prfC polypeptide is useful for screening for antibacterial

CC compounds which can be used as drugs. prfC polynucleotides can be used

CC in genetic immunisation (gene therapy) to protect against bacterial

CC infections. prfC polypeptides, polynucleotides and antagonists may be

CC used as a wound treatment to prevent adhesion of bacteria to matrix

CC proteins, as they interfere with the physical interaction between the

CC pathogen and mammalian host. prfC antibodies are also useful for inducing

CC an immune response to immunise and prevent disease, and for isolating

CC prfC clones or purifying the peptide by affinity chromatography. Diseases

CC diagnosed, prevented or treated include: otitis media, conjunctivitis,

CC pneumonia, bacteremia, sinusitis, pleural empyema, endocarditis and

CC especially meningitis.

Sequence 514 AA;

Query Match 39.48; Score 61; DB 39; Length 514;

Best Local Similarity 61.58; Pred. No. 3.85e+01;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db 21 agtltteqlylf 33
: | || |||||
1 TGARTINGOLLYF 13

RESULT 14

ID W03642 standard; peptide: 116 AA.

AC W03642;

DE 17-DEC-1996 (first entry)

KW Human cannabinoid GPR N-terminal sequence.

KW G-protein coupled receptor; ligand binding assay; transmembrane domain;

KW schizophrenia; dopamine; cAMP; adenosine; thrombin; adrenergic; opsin;

KW muscarinic acetylcholine; endothelin; bombesin; endocrine; rhodopsin;

Search completed: Tue Jan 11 15:46:47 2000
 Job time : 19 secs.

KW odorant; cytomegalovirus; serotonergic.
 OS Homo sapiens.
 PN US5508384-A.
 PD 16-APR-1996.
 PF 10-SEP-1992: 943236.
 PR 10-SEP-1992: US-943236.
 PR 09-SEP-1993: US-118270.
 PA (UYN) UNIV NEW YORK STATE.
 PI Murphy RB, Schuster DI;
 DR WPI: 96-208785/21.
 PT New dopamine receptor peptide - useful as antipsychotic agent, e.g.
 PT for treating schizophrenia
 PS Disclosure; Fig 8B(3); 184pp: English.
 CC Proteins W02657-W02720 represent a range of G-protein coupled receptor
 (GPR) proteins selected from cAMP, adenosine, muscarinic acetylcholine,
 adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
 odorant, cytomegaloviral and other GPR proteins. The peptides
 W03578-W03651 represent the N-terminal fragments of the above proteins.
 CC The receptor proteins were used to design polypeptides, pref. based on
 the transmembrane domains, for use in G-protein coupled receptor ligand
 binding assays. The polypeptide fragments retain biological activity
 CC such as binding a GPR ligand or modulating GPR ligand binding to a GPR
 (see W02747-W02999 for examples of polypeptide fragments).
 CC The polypeptide fragments can be used in compositions for treating
 CC subjects suffering from a pathology related to a GPR abnormality e.g. a
 CC psychotic disorder such as schizophrenia.
 SQ Sequence 116 AA;

Query Match

Best Local Similarity 38.7%; Score 60; DB 19; Length 116;
 Best Local Similarity 50.0%; Pred. No. 4.87e+01;

Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 14 rtttdillygsndiq 29
 |||:||||:|:|
 OY 4 RTINGQLLYFRANGVO 19

RESULT 15

ID R14196 standard; Protein; 472 AA.
 AC R14196;
 DT 19-DEC-1991 (first entry)
 DE Human cannabinoid receptor.
 KW Cannabis sativa; marijuana; drug test; substance K receptor.
 OS Homo sapiens.
 PN US7564075-A.
 PD 03-SEP-1991.
 PE 08-AUG-1990: 564075.
 PE 08-AUG-1990: US-564075.
 PI (USSH) NAT INST OF HEALTH.
 PI Matsuda L, Brownstein M, Bonner T;
 DR WPI: 91-303326/41.
 DR N-PDB: Q14003.
 PT DNA encoding mammalian cannabinoid receptor - used for producing
 PT receptor for screening drugs and ligands and in detection
 PS Disclosure; Fig 5; 25pp: English.
 CC SKR6 cDNA encoding the rat cannabinoid receptor was isolated from a
 CC rat cerebral cortex cDNA library. An EcoRI-XbaI fragment was used to
 CC screen a human cosmid library. A positive clone was identified and
 CC sequenced. There is ca. 97 per cent homology between this amino
 CC acid sequence deduced from the sequence of the positive human clone
 CC and rat cannabinoid receptor (see R14195). Recombinantly produced
 CC receptor can be used to screen for new drugs suitable for treatment
 CC of cannabinoid-treatable conditions, e.g. glaucoma, bronchial asthma,
 CC etc.
 SQ Sequence 472 AA;

Query Match

Best Local Similarity 38.7%; Score 60; DB 3; Length 472;
 Best Local Similarity 50.0%; Pred. No. 4.87e+01;

Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 14 rtttdillygsndiq 29
 |||:||||:|:|
 OY 4 RTINGQLLYFRANGVO 19

WILEY

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Jan 11 15:49:24 2000; MasPar time 1.77 Seconds
147.986 Million cell updates/sec
Lar output not generated.

Title: >US-09-290-049-3
Description: (1-22) from US09290049.pep
Perfect Score: 155
Sequence: 1 TGARTINGQLLYFRANGVQVKG 22

Scoring table: PAM 150
Gap 15

Searched: 122461 seqs, 11912985 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfilest1

Statistics: Mean 19.644; Variance 62.865; scale 0.312

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	114	73.5	22	1	US-08-057-	Sequence 3, Applicatio	2.93e+05
2	67	43.2	811	2	US-08-405-	Sequence 7, Applicatio	5.11e+00
3	67	43.2	811	1	US-08-480-	Sequence 29, Applicati	5.11e+00
4	67	43.2	812	1	US-08-480-	Sequence 29, Applicati	5.11e+00
5	67	43.2	2710	2	US-08-405-	Sequence 6, Applicatio	5.11e+00
6	67	43.2	2710	1	US-08-480-	Sequence 6, Applicatio	5.11e+00
7	63	40.6	404	2	US-08-666-	Sequence 7, Applicatio	1.32e+01
8	62	40.0	349	1	US-08-343-	Sequence 12, Applicati	1.67e+01
9	61	39.4	514	2	US-08-865-	Sequence 2, Applicatio	2.11e+01
10	59	38.1	91	1	US-08-480-	Sequence 8, Applicatio	3.36e+01
11	59	38.1	91	2	US-08-405-	Sequence 8, Applicatio	3.36e+01
12	59	38.1	453	2	US-08-485-	Sequence 8, Applicatio	3.36e+01
13	59	38.1	453	2	US-08-422-	Sequence 8, Applicatio	3.36e+01
14	59	38.1	453	2	US-08-476-	Sequence 6, Applicatio	3.36e+01
15	59	38.1	453	2	US-07-956-	Sequence 6, Applicatio	3.36e+01
16	59	38.1	453	2	US-08-611-	Sequence 8, Applicatio	3.36e+01
17	59	38.1	654	2	US-08-392-	Sequence 2, Applicatio	3.36e+01
18	59	38.1	998	3	PCT-US96-0	Sequence 2, Applicatio	3.36e+01
19	59	38.1	998	3	PCT-US95-0	Sequence 17, Applicati	4.32e+01
20	58	37.4	260	2	US-07-857-	Sequence 27, Applicati	5.32e+01
21	57	36.8	363	4	5223606-6	Patent No. 5223606.	5.32e+01
22	57	36.8	488	1	US-08-001-	Sequence 2, Applicatio	5.32e+01
23	57	36.8	488	1	US-07-794-	Sequence 2, Applicatio	5.32e+01

RESULT ID	1	STANDARD	PRT	22 AA.
24	57	36.8	492	1 US-07-794- Sequence 4, Applicatio
25	57	36.8	492	1 US-08-001- Sequence 4, Applicatio
26	57	36.8	501	4 5168064-4 Patent No. 5168064.
27	57	36.8	501	1 US-08-271- Sequence 2, Applicatio
28	57	36.8	501	1 US-07-687- Sequence 2, Applicatio
29	57	36.8	501	1 US-08-434- Sequence 2, Applicatio
30	56	36.1	15	4 5439796-5 Patent No. 5439796.
31	56	36.1	54	2 US-08-797- Sequence 9, Applicatio
32	56	36.1	448	2 US-08-231- Sequence 23, Applicatio
33	56	36.1	506	2 US-08-635- Sequence 2, Applicatio
34	55	35.5	265	2 US-09-002- Sequence 2, Applicatio
35	55	35.5	265	1 US-08-461- Sequence 2, Applicatio
36	55	35.5	265	1 PCT-US94-1 Sequence 3, Applicatio
37	55	35.5	274	1 US-08-287- Sequence 3, Applicatio
38	55	35.5	448	1 US-08-074- Sequence 3, Applicatio
39	55	35.5	448	3 PCT-US94-0 Sequence 3, Applicatio
40	55	35.5	971	1 US-08-446- Sequence 19, Applicati
41	55	35.5	971	1 US-09-066- Sequence 19, Applicati
42	55	35.5	971	1 US-08-446- Sequence 19, Applicati
43	55	35.5	971	2 US-08-805- Sequence 19, Applicati
44	55	35.5	971	2 US-08-064- Sequence 19, Applicati
45	55	35.5	993	1 US-08-446- Sequence 25, Applicati

Sequence 3, Application US/08057162B

Sequence 3, Application US/08057162B

Patent No. 5686075

GENERAL INFORMATION:

APPLICANT: Taubman, Martin A.

TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINE FOR DENTAL CARRIES

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: Two Millitia Drive

CITY: Lexington

STATE: Massachusetts

COUNTRY: USA

ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/057.162B

FILING DATE: 30-APR-1993

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/877.295

FILING DATE: 01-MAY-1992

ATTORNEY/AGENT INFORMATION:

NAME: Wagner, Richard W.

REGISTRATION NUMBER: 34,480

REFERENCE/DOCKET NUMBER: FDC92-01A

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 861-6240

TELEFAX: (617) 861-9540

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 22 amino acids

TYPE: amino acid

STRANDEDNESS: single

CC TOPOLOGY: 1linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 22 AA; 2366 MW; 2475 CN;
Query Match 73.5%; Score 114; DB 1; Length 22;
Best Local Similarity 77.3%; Pred. No. 2.99e-05;
Matches 17; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
DB 1 TGAQTIGKGLYFRANGOVKG 22
1 TGARTINGQLLYFRANGOVKG 22
OY
RESULT 2 STANDARD; PRT; 811 AA.
ID US-08-405-496A-7
XX
AC xxxxxx
DE
Sequence 7, Application US/08405496A
CC Sequence 7, Application US/08405496A
CC Patent No. 5919665
CC GENERAL INFORMATION:
CC APPLICANT: WILLIAMS, JAMES A.
CC TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
CC TITLE OF INVENTION: NEUROTOXIN
CC NUMBER OF SEQUENCES: 30
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MEDLEN & CARROLL, LLP
CC STREET: 220 MONTGOMERY STREET, SUITE 2200
CC CITY: SAN FRANCISCO
CC STATE: CALIFORNIA
CC COUNTRY: USA
CC ZIP: 94104
CC
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/405,496A
CC FILING DATE: 16-MAR-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/329,154
CC FILING DATE: 25-OCT-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/161,907
CC FILING DATE: 02-DEC-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/985,321
CC FILING DATE: 04-DEC-1992
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/429,791
CC FILING DATE: 31-OCT-1989
CC ATTORNEY/AGENT INFORMATION:
CC NAME: INGOLIA, DIANE E.
CC REGISTRATION NUMBER: 40,027
CC REFERENCE/DOCKET NUMBER: OPHD-01308
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 705-8410
CC TELEFAX: (415) 397-8338
CC INFORMATION FOR SEQ ID NO: 7:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 811 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC MOLECULE TYPE: protein
CC SEQUENCE 811 AA; 91921 MW; 3356332 CN;

Query Match 43.2%; Score 67; DB 2; Length 811;
Best Local Similarity 52.2%; Pred. No. 5.11e-00;
Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;
DB 587 TGLRTIDGKKYYNTAVAVTG 609
1 TGARTINGQLLYFRAN-GVOVKG 22
OY
RESULT 3 STANDARD; PRT; 811 AA.
ID US-08-480-604A-7
XX
AC xxxxxx
DE
Sequence 7, Application US/08480604A
CC Sequence 7, Application US/08480604A
CC Patent No. 5736139
CC GENERAL INFORMATION:
CC APPLICANT: KINK, JOHN A.
CC APPLICANT: THALLEY, BRUCE S.
CC APPLICANT: PADRYE, NISRA V.
CC APPLICANT: FTRCA, JOSEPH R.
CC APPLICANT: STAFFORD, DOUGLAS C.
CC TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
CC TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
CC NUMBER OF SEQUENCES: 32
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MEDLEN & CARROLL, LLP
CC STREET: 220 MONTGOMERY STREET, SUITE 2200
CC CITY: SAN FRANCISCO
CC STATE: CALIFORNIA
CC COUNTRY: UNITED STATES OF AMERICA
CC ZIP: 94104
CC
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/480,604A
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/422,711
CC FILING DATE: 14-APR-1995
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/405,496
CC FILING DATE: 16-MAR-1995
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/329,154
CC FILING DATE: 25-OCT-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/161,907
CC FILING DATE: 02-DEC-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/985,321
CC FILING DATE: 04-DEC-1992
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/429,791
CC FILING DATE: 31-OCT-1989
CC ATTORNEY/AGENT INFORMATION:
CC NAME: INGOLIA, DIANE E.
CC REGISTRATION NUMBER: 40,027
CC REFERENCE/DOCKET NUMBER: OPHD-01763
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 705-8410
CC TELEFAX: (415) 397-8338
CC INFORMATION FOR SEQ ID NO: 7:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 811 amino acids

CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC MOLECULE TYPE: protein
CC SEQUENCE 811 AA; 91921 MW; 3356332 CN;
SQ
Query Match 43.28; Score 67; DB 1; Length 811;
Best Local Similarity 52.28; Pred. No. 5.11e+00;
Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;
Db 587 TGLRTIDGKKRYENTNAYVATG 609
QY 1 TGARTINGOLLYFRAN-GVOYKG 22
RESULT 4
ID US-08-480-604A-29 STANDARD; PRT; 812 AA.
XX
AC xxxxxx
XX
Sequence 29, Application US/08480604A
XX Patent No. 5736139
CC GENERAL INFORMATION:
CC APPLICANT: KINK, JOHN A.
CC APPLICANT: THALLEY, BRUCE S.
CC APPLICANT: PADHAY, NISHA V.
CC APPLICANT: FIRCA, JOSEPH R.
CC APPLICANT: STAFFORD, DOUGLAS C.
CC TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
CC PREVENTION OF C. DIFFICILE DISEASE
CC NUMBER OF SEQUENCES: 32
CC CORESPONDENCE ADDRESS:
CC ADDRESSEE: MEDLEN & CARROLL, LLP
CC STREET: 220 MONTGOMERY STREET, SUITE 2200
CC CITY: SAN FRANCISCO
CC STATE: CALIFORNIA
CC COUNTRY: UNITED STATES OF AMERICA
CC ZIP: 94104
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/480,604A
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/422,711
CC FILING DATE: 14-APR-1995
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/405,496
CC FILING DATE: 16-MAR-1995
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/329,154
CC FILING DATE: 25-OCT-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/161,907
CC FILING DATE: 02-DEC-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/985,321
CC FILING DATE: 04-DEC-1992
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/429,791
CC FILING DATE: 31-OCT-1989
CC ATTORNEY/AGENT INFORMATION:
CC NAME: INGOLIA, DIANE E.
CC REGISTRATION NUMBER: 40,027
CC REFERENCE/DOCKET NUMBER: OPHD-01763

CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 705-8410
CC TELEFAX: (415) 397-8338
CC INFORMATION FOR SEQ ID NO: 29:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 812 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 812 AA; 92022 MW; 3364670 CN;
SQ
Query Match 43.28; Score 67; DB 1; Length 812;
Best Local Similarity 52.28; Pred. No. 5.11e+00;
Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;
Db 588 TGLRTIDGKKRYENTNAYVATG 610
QY 1 TGARTINGOLLYFRAN-GVOYKG 22
RESULT 5
ID US-08-405-496A-6 STANDARD; PRT; 2710 AA.
XX
AC xxxxxx
XX
Sequence 6, Application US/08405496A
XX Patent No. 5919665
CC GENERAL INFORMATION:
CC APPLICANT: WILLIAMS, JAMES A.
CC TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
CC NUMBER OF SEQUENCES: 30
CC CORESPONDENCE ADDRESS:
CC ADDRESSEE: MEDLEN & CARROLL, LLP
CC STREET: 220 MONTGOMERY STREET, SUITE 2200
CC CITY: SAN FRANCISCO
CC STATE: CALIFORNIA
CC COUNTRY: USA
CC ZIP: 94104
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/405,496A
CC FILING DATE: 16-MAR-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/329,154
CC FILING DATE: 25-OCT-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/161,907
CC FILING DATE: 02-DEC-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/985,321
CC FILING DATE: 04-DEC-1992
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/429,791
CC FILING DATE: 31-OCT-1989
CC ATTORNEY/AGENT INFORMATION:
CC NAME: INGOLIA, DIANE E.
CC REGISTRATION NUMBER: 40,027
CC REFERENCE/DOCKET NUMBER: OPHD-01308
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 705-8410
CC TELEFAX: (415) 397-8338
CC INFORMATION FOR SEQ ID NO: 6:

CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 2710 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 2710 AA: 308053 MW: 38468929 CN;
DB 2460 TGLRTIDGKKYFNTNVAVTG 2482
QY 1 TGARTINGQLIFRAN-GVOYKG 22
Query Match 43.28; Score 67; DB 2; Length 2710;
Best Local Similarity 52.28; Pred. No. 5.11e+00;
Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;
RESULT 6 STANDARD; PRT: 2710 AA.
ID US-08-480-604A-6
MA xxxxxx
DE Sequence 6, Application US/08480604A
XX Patent No. 5736139
CC GENERAL INFORMATION:
CC APPLICANT: KIRK, JOHN A.
CC APPLICANT: THALLEY, BRUCE S.
CC APPLICANT: PADHYE, NISHA V.
CC APPLICANT: FIRCA, JOSEPH R.
CC APPLICANT: STAFFORD, DOUGLAS C.
CC TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
CC TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
CC NUMBER OF SEQUENCES: 32
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MEDLEN & CARROLL, LLP
CC STREET: 220 MONTGOMERY STREET, SUITE 2200
CC CITY: SAN FRANCISCO
CC STATE: CALIFORNIA
CC COUNTRY: UNITED STATES OF AMERICA
CC ZIP: 94104
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/480,604A
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/422,711
CC FILING DATE: 14-APR-1995
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/405,496
CC FILING DATE: 16-MAR-1995
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/329,154
CC FILING DATE: 25-OCT-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/161,907
CC FILING DATE: 02-DEC-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/985,321
CC FILING DATE: 04-DEC-1992
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/429,791
CC FILING DATE: 31-OCT-1989
CC ATTORNEY/AGENT INFORMATION:
CC NAME: INGOLIA, DIANE E.
CC REGISTRATION NUMBER: 40,027

CC REFERENCE/DOCKET NUMBER: OPFD-01763
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 705-8410
CC TELEFAX: (415) 397-8338
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 2710 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 2710 AA: 308053 MW: 38468929 CN;
DB 2460 TGLRTIDGKKYFNTNVAVTG 2482
QY 1 TGARTINGQLIFRAN-GVOYKG 22
Query Match 43.28; Score 67; DB 1; Length 2710;
Best Local Similarity 52.28; Pred. No. 5.11e+00;
Matches 12; Conservative 3; Mismatches 7; Indels 1; Gaps 1;
RESULT 7 STANDARD; PRT: 404 AA.
ID US-08-666-367B-7
MA xxxxxx
DE Sequence 7, Application US/08666367B
XX Patent No. 5854042
CC GENERAL INFORMATION:
CC APPLICANT: Shuichi TSUJI et al.
CC TITLE OF INVENTION: NOVEL SUGAR-CHAIN SYNTHETASE AND PROCESS FOR
CC TITLE OF INVENTION: PRODUCING THE SAME
CC NUMBER OF SEQUENCES: 8
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Wenderoth, Lind & Ponack
CC STREET: 805 Fifteenth Street, N.W., #700
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: U.S.A.
CC ZIP: 20005
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
CC COMPUTER: IBM compatible
CC OPERATING SYSTEM: MS-DOS
CC SOFTWARE: Wordperfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/666,367B
CC FILING DATE: August 19, 1996
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Warren M. Cheek, Jr.
CC REGISTRATION NUMBER: 33,367
CC REFERENCE/DOCKET NUMBER:
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-371-8850
CC TELEFAX:
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 7:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 404 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: G. gallus (chicken)
CC SEQUENCE 404 AA: 45826 MW: 844143 CN;
SQ

CC APPLICANT: THALLEY, BRUCE S.
CC APPLICANT: PADHYE, NISHA V.
CC APPLICANT: FIRCA, JOSEPH R.
CC APPLICANT: STAFFORD, DOUGLAS C.
CC TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
CC TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
CC NUMBER OF SEQUENCES: 32
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MEDLEN & CARROLL, LLP
CC STREET: 220 MONTGOMERY STREET, SUITE 2200
CC CITY: SAN FRANCISCO
CC STATE: CALIFORNIA
CC COUNTRY: UNITED STATES OF AMERICA
CC ZIP: 94104
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/480,604A
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/422,711
CC FILING DATE: 14-APR-1995
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/405,496
CC FILING DATE: 16-MAR-1995
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/329,154
CC FILING DATE: 25-OCT-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/161,907
CC FILING DATE: 02-DEC-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/985,321
CC FILING DATE: 04-DEC-1992
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/429,791
CC FILING DATE: 31-OCT-1989
CC TELECOMMUNICATION INFORMATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: INGOLIA, DIANE E.
CC REGISTRATION NUMBER: 40,027
CC REFERENCE/DOCKET NUMBER: OPHD-01763
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 705-8410
CC TELEFAX: (415) 397-8338
CC INFORMATION FOR SEQ ID NO: 8:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 91 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC MOLECULE TYPE: protein
CC SEQUENCE 91 AA: 10432 MW: 39683 CN:
SQ
Query Match 38.1%; Score 59; DB 1; Length 91;
Best Local Similarity 46.7%; Pred. No. 3.36e+01;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
Db 3 KIINGKHFFYNNDGV 17
QY 4 RTINGQLTYFRANGV 18
RESULT 11
ID US-08-405-496A-8 STANDARD; PRT; 91 AA.
AC xxxxxx
XX
DT
XX

DE Sequence 8, Application US/08405496A
XX Sequence 8, Application US/08405496A
CC Patent No. 5919665
CC GENERAL INFORMATION:
CC APPLICANT: WILLIAMS, JAMES A.
CC TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
CC TITLE OF INVENTION: NEUROTOXIN
CC NUMBER OF SEQUENCES: 30
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MEDLEN & CARROLL, LLP
CC STREET: 220 MONTGOMERY STREET, SUITE 2200
CC CITY: SAN FRANCISCO
CC STATE: CALIFORNIA
CC COUNTRY: USA
CC ZIP: 94104
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/405,496A
CC FILING DATE: 16-MAR-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/329,154
CC FILING DATE: 25-OCT-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/161,907
CC FILING DATE: 02-DEC-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/985,321
CC FILING DATE: 04-DEC-1992
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/429,791
CC FILING DATE: 31-OCT-1989
CC ATTORNEY/AGENT INFORMATION:
CC NAME: INGOLIA, DIANE E.
CC REGISTRATION NUMBER: 40,027
CC REFERENCE/DOCKET NUMBER: OPHD-01308
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 705-8410
CC TELEFAX: (415) 397-8338
CC INFORMATION FOR SEQ ID NO: 8:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 91 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC MOLECULE TYPE: protein
CC SEQUENCE 91 AA: 10432 MW: 39683 CN:
SQ
Query Match 38.1%; Score 59; DB 2; Length 91;
Best Local Similarity 46.7%; Pred. No. 3.36e+01;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
Db 3 KIINGKHFFYNNDGV 17
QY 4 RTINGQLTYFRANGV 18
RESULT 12
ID US-08-485-607-6 STANDARD; PRT; 453 AA.
AC xxxxxx
XX
DT
XX
DE Sequence 6, Application US/08485607
CC Sequence 6, Application US/08485607
CC Patent No. 5792627

CC REGISTRATION NUMBER: 33,268
 CC REFERENCE/DOCKET NUMBER: ARCD:058
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: 1-312-744-0090
 CC TELEFAX: 1-312-755-4489
 CC INFORMATION FOR SEQ ID NO: 6:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 453 amino acids
 CC TYPE: Amino acid
 CC STRANDEDNESS: Single
 CC TOPOLOGY: Linear
 CC MOLECULE TYPE: Peptide
 CC SEQUENCE 453 AA; 49608 MW; 983857 CN;

Query Match 38.1%; Score 59; DB 2; Length 453;
 Best Local Similarity 36.8%; Pred. No. 3.36e+01;
 Matches 7; Conservative 4; Mismatches 8; Indels 0; Gaps 0;
 315 RIAOGEALRFRQADIDLRG 333
 QY 4 RTINGOLLYFRANGVOYKG 22

RESULT 15
 ID US-07-956-700B-6 STANDARD; PRT: 453 AA.
 AC xxxxxx
 XX
 XX
 DT
 XX
 XX

Sequence 6, Application US/07956700B

CC Sequence 6, Application US/07956700B
 CC Patent No. 5539092
 CC GENERAL INFORMATION:
 CC APPLICANT: Robert Haselkorn and Piotr Gornicki
 CC TITLE OF INVENTION: Cyanobacterial and Plant Acetyl-CoA
 CC NUMBER OF SEQUENCES: 116
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: Arnold, White & Durkee
 CC STREET: 321 No. 5539092th Clark Street
 CC CITY: Chicago
 CC STATE: Illinois
 CC COUNTRY: USA
 CC ZIP: 60610
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy Disk
 CC COMPUTER: IBM PC Compatibld
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: ASCII-DOS
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/07/956,700B
 CC FILING DATE: 19921002
 CC CLASSIFICATION: 800
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Thomas E. No. 5539092thrup
 CC REGISTRATION NUMBER: 33,268
 CC REFERENCE/DOCKET NUMBER: ARCD:058
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: 1-312-744-0090
 CC TELEFAX: 1-312-755-4489
 CC INFORMATION FOR SEQ ID NO: 6:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 453 amino acids
 CC TYPE: Amino acid
 CC STRANDEDNESS: Single
 CC TOPOLOGY: Linear
 CC MOLECULE TYPE: Peptide
 CC SEQUENCE 453 AA; 49608 MW; 983857 CN;

Query Match 38.1%; Score 59; DB 1; Length 453;
 Best Local Similarity 36.8%; Pred. No. 3.36e+01;

Matches 7; Conservative 4; Mismatches 8; Indels 0; Gaps 0;
 Db 315 RIAOGEALRFRQADIDLRG 333
 QY 4 RTINGOLLYFRANGVOYKG 22

Search completed: Tue Jan 11 15:49:31 2000
 Job time : 7 secs.

[illegible]

Release 3.1A John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Jan 11 15:41:17 2000; MasPar time 3.90 Seconds
109.105 Million cell updates/sec
T1 output not generated.

```

```
Title: >US-09-290-049-2
Description: (1-20) from US09290049.pep
Perfect Score: 135
Sequence: 1 VPSYSFIRTAHDSVQDLIA 20
```

Scoring table: PAM 150
Gap 15

Searched: 170751 seqs, 21266608 residues

```
post-processing: Minimum Match 0%
Listing first 45 summaries
```

```
Database:
a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39
```

Statistics: Mean 20.274; Variance 61.611; scale 0.3299

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB	ID	Description	pred. No
1	97	71.9	1592	6	R32925	Glucosyltransferase	2.44e+03
2	71	52.6	1577	16	R91047	Alpha-D-glucosyltransferase	2.33e+00
3	63	47.7	825	4	R22237	Sequence of the "gs"	1.71e+01
4	63	46.7	873	14	R79459	Infectious Laryngotracheitis	1.71e+01
5	45.2	52.2	344	30	W52812	Human induced tumour	2.77e+01
6	61	45.2	579	16	R86406	Human matrix metalloproteinase	2.77e+01
7	61	45.2	582	16	R86407	Human matrix metalloproteinase	2.77e+01
8	61	45.2	582	14	R75648	Human matrix metalloproteinase	2.77e+01
9	60	44.4	146	6	R32010	Human placenta derived	2.77e+01
10	60	44.4	146	6	R32010	Rp15-TIA-1.	3.53e+01
11	60	44.4	156	8	R12539	Cytotoxic T lymphocyte	3.53e+01
12	60	44.4	156	8	R40840	Cytotoxic T lymphocyte	3.53e+01
13	60	44.4	375	6	R32011	Glycocalyx derived	3.53e+01
14	60	44.4	375	6	R32009	TAR.	3.53e+01
15	59	43.7	465	31	W54351	Rp40-TIA-1.	3.53e+01
16	58	43.0	557	24	W24096	Vimentin.	4.48e+01
			159	10	R33249	Enzyme for production of	5.69e+01
						Escherichia coli	

17	57	42.2	1010	18	R91823	Human immunodeficient	7.21e+011
18	56	41.5	133	33	W61232	Streptococcus pneumoniae	9.12e+010
19	56	41.5	1014	35	W66474	HIV-1 strain YBF30 po	9.12e+010
20	55	40.7	326	11	RS7080	Fasciola hepatica cat	1.15e+020
21	55	40.7	792	11	R55602	X-prolyl-diethylglycyl-a	1.15e+020
22	55	40.7	1002	37	W72993	HIV isolate LAV.MAL p	1.15e+020
23	55	40.7	1002	1	P81861	Sequence encoded by L	1.15e+020
24	55	40.7	1003	2	R08060	HIV-1 pol protein of	1.15e+020
25	54	40.0	244	21	W14444	Carb gene product.	1.45e+020
26	54	40.0	244	26	R64175	Carbapenem R.	1.45e+020
27	54	40.0	244	26	R64176	Carbapenem R.	1.45e+020
28	54	40.0	432	23	W14081	S.thermophilus exopol	1.45e+020
29	54	40.0	473	23	W22180	S.thermophilus exopol	1.45e+020
30	54	40.0	525	9	R47467	Rinderpest virus nuc	1.45e+020
31	54	40.0	582	30	W52134	Rabbit membrane-type	1.45e+020
32	54	40.0	623	6	R31522	Carrot reverse transc	1.45e+020
33	54	40.0	775	33	W63748	Human semaphorin.	1.45e+020
34	54	40.0	845	25	W26475	KSHV glycoprotein B.	1.45e+020
35	54	40.0	1068	8	R43342	Human p110.	1.45e+020
36	54	40.0	1068	9	R43341	p110.	1.45e+020
37	54	40.0	1068	9	R46294	PtdIns 3-kinase 110 k	1.45e+020
38	54	40.0	1076	31	W53360	Hypo-thetical protein	1.45e+020
39	53	39.3	12	16	R83328	Kb-binding random pep	1.83e+020
40	53	39.3	285	37	W74879	Human secreted protei	1.83e+020
41	53	39.3	395	6	R29520	Bt toxin 69D1.	1.83e+020
42	53	39.3	395	6	R28813	Bt toxin 69D1.	1.83e+020
43	53	39.3	934	5	W20016	Sequence of p20, VP4,	1.83e+020
44	52	38.5	712	30	W53110	POL protein contained	2.30e+020
45	52	38.5	1050	39	W67642	A serine/threonine pr	2.30e+020

ALIGNMENTS

```

RESULT      1
ID          R32925 standard; Protein; 1592 AA.
AC          R32925.
DT          28-JUN-1993 (first entry)
DE          Glucosyltransferase I.
KW          GT-1; Streptococcus; dental; caries.
OS          Streptococcus sobrinus.
PN          J05023188-A.
PD          02-FEB-1993.
PF          25-JUL-1991; 186592.
PR          25-JUL-1991; JP-186592.
PA          (FUKU/) FUKUI I.
RA          (KATO/) KATO K.
DR          WPI; 93-079449/10.
DR          N-PSDB; Q37760.
PT          DNA sequence glucosyl:transferase-I - comprises Streptococcus
PT          sobrinus DNA sequence with at least one nucleotide added or
PT          deleted
PS          Claim 13; Page 15; 29pp; Japanese.
CC          The DNA sequence from Streptococcus sobrinus strain 6715 encodes
CC          glucosyltransferase-I (and mutants). The DNA was obt'd. by treating
CC          S. sobrinus 6715 with mutanolysin, extracting the chromosomal DNA,
CC          partially digesting with SauHI and fractionating on agarose gel.
CC          The 3-5 kbp fragment was ligated into pUC18 and E. coli JM109
CC          transformed with it. A GT-1 expressing clone was isolated and
CC          sequenced. The clone may be used in the development of a drug for
CC          dental caries.
SQ          Sequence 1592 AA;

      Query Match      71.9%; Score 97; DB 6; Length 1592;
      Best Local Similarity 84.2%; Pred. NO. 2.44e-03;
      Matches 16; Conservative 1; Mismatches 1; Indels 1; Gaps 1

Db      548 vpsysfar-ahdsevgd1 565
Oy      1 vpsysfirtahdsevgd1 19

```

AC R91047: 22-MAY-1996 (first entry)
 DT Alpha-D-glucosyltransferase.
 DE Alpha-D-glucosyltransferase.
 KW sucrose; transgenic plant; cloning; *Escherichia coli*;
 KW phage lambda-Cl3 vector; plasmid pSGS502;
 KW gene transfer; crop improvement; storage carbohydrate; pasture;
 KW feedstuff; senescence; dextran; binder; food; pharmaceutical.
 OS *Streptococcus salivarius* strain ATCC 25975.
 PN MO9606173-A1.
 PD 29-FEB-1996.
 PE 24-AUG-1995: AU00527.
 PR 24-AUG-1994: AU-007443.
 PA (GIEF/) GIEFARD P M.
 PA (JACO/) JACQUES N A.
 PA (STMP/) STIMPSON C L.
 FI Giffard PM, Jacques NA, Simpson CL;
 OR WP1: 96-151376/15.

Plants cont. new bacterial DNA encoding glucosyl transferase activity - retain higher levels of stored carbohydrate(s) in a form readily digestible by ruminants
PR Claim 4: Page 16-20: 31pp: English.
The sequence represents an alpha-D-glucosyltransferase from *Streptococcus salivarius*. The enzyme is primer-independent, and produces soluble glucan from sucrose. A gene encoding the enzyme may be cloned and expressed in *Escherichia coli* using a subclone of phage lambda-CJ3, e.g. plasmid pGS6501 or plasmid pGS6502. The DNA may also be expressed in a transgenic plant, to improve the level of stored carbohydrate in a pasture plant which normally contains low levels, or to prevent degradation of stored carbohydrate during plant senescence. Dextran may be isolated from the plant, for use as a food binder or pharmaceutical additive. primer independence ensures that the enzyme will be functional in plants. The glucan is poorly degraded in plants but easily degraded by bacteria in the rumen of grazing livestock.
Sequence 1577 AA:
90

Query Match	52.68;	Score 71;	DB 16;	Length 1577;
Best Local Similarity	57.98;	Pred. No. 2.33e+00;		
Matches	11;	Conservative	5;	Mismatches 2; Indels 1; Gaps 1

```
Db      660 anylfvr-ahdsevgavla 67
      ::| |::| | | | | | | ::|
QY      2 PSYSFIRTAHDSEVQDLIA 20
```

FRONT 3
R22237 standard; Protein; 825 AA
P33337.

DT 23-JUL-1992 (first entry)
DE Sequence of the "gb" homologue of ILTV encoded by the EcoRI "u" (unique) genome fragment

KW Subunit vaccine; immunogen; glycoprotein; promoter
OS Infectious laryngotracheitis virus.

	peptide	1.16 /label-signal
FT		
FT		

PD	05-MAR-1992
PF	23-AUG-1991

PA (WEBS-) WEBSTER A PTY LTD.
PI Sheppard MG, Pridaux C, Johnson M, Fahey KJ, York JJ,

DR WPI; 92-096898/12.
DR N-PSDB; Q22989.

PT administered by aerosol
PS Example: Fig 11; 122pp; English.

CC The inventors claim a non-infectious subunit vaccine for use against
CC ILTV which comprises a glycoprotein of ILTV. They also claim DNA

CC encoding the vaccine, a synthetic polypeptide displaying the
CC antigenicity of the 205k complex or 60k ILTV glycoprotein and a
CC recombinant ILTV with heterologous DNA inserted into a non-essential
CC region of the genome; and a recombinant DNA mol. comprising an ILTV
CC promoter region operatively linked to a heterologous DNA sequence.
CC The promoter region is the ILTV gp60 promoter. gp205 (gpB) promoter
CC or the ORF3 promoter region. KpnI/ORF3 is located 5' of the gp60
CC gene. 825 Aa:
SQ Sequence

Query Match	46.7%;	Score 63;	DB 4;	Length 825;
Best Local Similarity	33.3%;	Pred. NO. 1.71e+01;		
Matches	5;	Conservative	7;	Mismatches 3;
				Indels 0;
				Gaps 0;

```
Db 570 yltivirgnaseleev 584  
|:|:| :|:|:  
QY 4 XSFIRTAHDSEVQDL 18
```

RESULT 4
ID R79459 standard; Protein; 873 AA
AC R79459.

DT	16-JAN-1996	(first entry)
DE	Infectious Laryngotracheitis	
RW	Infectious Laryngotracheitis	

KW vaccine; immunosay; detection; identification; avipox; chicken,
fowl; probe; antisense; gene expression.
KW Callid homocytine 1

	Key	Location/Qualifiers
FH	misc_difference	656..659
FT		/notes #these am ino

FT		decoded polypeptide sequence given in the
FT		specification."
DN	nsc442931-7	

PD	22-AUG-1995
PF	29-OCT-1991
DP	28-OCT-1991

PR 19-NOV-1993; US-156866
PA (UYDE) UNIV DELAWARE.
PI 19-NOV-1993; US-156866

DR WPI; 95-302091/39.
DR N-PSDB; Q97351.
DE T001304 Tefest40.

PT develop prods. for detection and for vaccine(s), partic. for immunising fowl.

CC The infectious laryngotracheitis virus (ILV) gB glycoprotein
CC nucleic acids can be used as probes, for antisense control o

CC polypeptides can be used for the production of antibodies (both
CC used in immunossays) and in vaccines. Recombinant avipox virus

CC chickens. 873 AA;
SQ Sequence

Query Match	46.7%;	Score 63;	DB 14;	Length 873;
Best Local Similarity	33.3%;	Pred. No. 1.71e+01;		
Matches	5;	Conservative	7;	Mismatches 3; Indels 0; Gaps 0;

```

Db      618 ytfvrqnaseieev 632
      1:1:1 : 11:::
QY      4 YSFIRTAHDSEVQDL 18

```

RESULT	5
ID	W52812 standard; Protein; 344 AA

DE Human induced tumour protein.
KW Human induced tumour protein; HTP; treatment; prevention; cancer;
??

OS FH	Homo sapiens. Key	Location/Qualifiers
----------	----------------------	---------------------

FT Misc_difference 284 /note="encoded by CTR"
 PN WO9806846-A1.
 PD 19-FEB-1998.
 PF 14-AUG-1997: U14378.
 PR 16-AUG-1996: US-689974.
 PA (INCY-) INCYTE PHARM INC.
 PI Au-Young J, Hawkins PR;
 PI WPI: 98-159538/14.
 DR N-PSDB: V20913.
 PT Human tumour induced protein and related nucleic acid, vectors -
 PT transformed cells and antibodies, useful for inducing and monitoring of
 PT differentiation of cancer cells, and for diagnosis and monitoring of
 PT treatment
 PS Claim 1; Fig 1; 61pp; English.
 CC The present sequence is human induced tumour protein (HITP),
 CC which can be used to treat and prevent cancer, specifically
 CC melanoma and carcinoma of the breast, colon or brain. HITP induces
 CC cancer cell differentiation, and stops cell division.
 SQ Sequence 344 AA;
 DB 221 ysisrtnsqvynk1 235
 QY 4 YSFIRTAHSEVQDL 18
 RESULT 6 45.2%; Score 61; DB 30; Length 344;
 ID R86406 standard; Protein: 579 AA.
 AC R86406;
 DT 15-MAY-1996 (first entry)
 DE Human matrix metalloproteinase MMP1a.
 KM Human; matrix metalloproteinase; MMP; extracellular matrix; inhibitor;
 KM activator; zinc-binding region.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT region 88..94
 FT /note="conserved region within pro-sequence,
 FT contains a Cys residue"
 FT region 105..110
 FT /note="conserved region within pro-sequence"
 FT domain 236..246
 FT /label="catalytic
 FT /note="contains 3 His residues"
 FT region 536..556
 FT /label="hydrophobic
 FT WO9525171-A2.
 DT 21-SEP-1995.
 PR 17-MAR-1995: D00357.
 PR 17-MAR-1994; DE-409663.
 PR 21-OCT-1994; DE-438838.
 PA (DELB-) DELBROECK CENT MOLEKULARE MEDIZIN MAX.
 PI Hinzmann B, Will H;
 PI WPI: 95-336975/43.
 DR N-PSDB: T03436.
 PT New matrix metallo:proteinase(s) and DNA encoding them - also vectors,
 PT recombinant cells and complexes with ligands, useful as proteolytic
 PT agents and for identification of specific inhibitors and activators
 PS Claim 8; Page 30; 85pp; German.
 CC The present sequence is that of human matrix metalloproteinase
 CC MMP1a. The protein has mol. wt. 65901 and comprises a signal
 CC peptide, pro-region, conserved catalytic domain and haemopexin-
 CC homologous sequence, all typical of known metalloproteinases. MMP1a
 CC differs from known metalloproteinases in having a hydrophobic region
 CC near the C-terminus, suggesting that the protein is membrane-bound.
 CC MMP proteins can hydrolyse extracellular matrix proteins and are
 CC involved in certain pathologies. The new MMP will be useful for
 CC detecting MMP inhibitors or activators and to generate diagnostic
 CC antibodies.
 SQ Sequence 579 AA;

Query Match 45.2%; Score 61; DB 16; Length 579;
 Best Local Similarity 29.4%; Pred. No. 2,77e+01;
 Matches 5; Conservative 8; Mismatches 4; Indels 0; Gaps 0;
 DB 160 pyatireghekqadml 176
 QY 3 YSFIRTAHSEVQDL 19
 RESULT 7
 ID R86407 standard; Protein: 582 AA.
 AC R86407;
 DT 15-MAY-1996 (first entry)
 DE Human matrix metalloproteinase MMP1b.
 KM Human; matrix metalloproteinase; MMP; extracellular matrix; inhibitor;
 KM activator; zinc-binding region.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT region 91..97
 FT /note="conserved region within pro-sequence,
 FT contains a Cys residue"
 FT region 108..113
 FT /note="conserved region within pro-sequence"
 FT domain 239..249
 FT /label="catalytic
 FT /note="contains 3 His residues"
 FT region 539..559
 FT /label="hydrophobic
 FT WO9525171-A2.
 DT 21-SEP-1995.
 PR 17-MAR-1995: D00357.
 PR 17-MAR-1994; DE-409663.
 PR 21-OCT-1994; DE-438838.
 PA (DELB-) DELBROECK CENT MOLEKULARE MEDIZIN MAX.
 PI Hinzmann B, Will H;
 PI WPI: 95-336975/43.
 DR N-PSDB: T03437.
 PT New matrix metallo:proteinase(s) and DNA encoding them - also vectors,
 PT recombinant cells and complexes with ligands, useful as proteolytic
 PT agents and for identification of specific inhibitors and activators
 PS Claim 9; Page 30-31; 85pp; German.
 CC The present sequence is that of human matrix metalloproteinase
 CC MMP1b. The protein has mol. wt. 65900 and comprises a signal
 CC peptide, pro-region, conserved catalytic domain and haemopexin-
 CC homologous sequence, all typical of known metalloproteinases. MMP1b
 CC differs from known metalloproteinases in having a hydrophobic region
 CC near the C-terminus, suggesting that the protein is membrane-bound.
 CC MMP proteins can hydrolyse extracellular matrix proteins and are
 CC involved in certain pathologies. The new MMP will be useful for
 CC detecting MMP inhibitors or activators and to generate diagnostic
 CC antibodies.
 SQ Sequence 582 AA;
 Query Match 45.2%; Score 61; DB 16; Length 582;
 Best Local Similarity 29.4%; Pred. No. 2,77e+01;
 Matches 5; Conservative 8; Mismatches 4; Indels 0; Gaps 0;
 DB 163 pyatireghekqadml 179
 QY 3 YSFIRTAHSEVQDL 19
 RESULT 8
 ID R75648 standard; Protein: 582 AA.
 AC R75648;
 DT 31-JAN-1996 (first entry)
 DE Human placenta derived metalloproteinase.
 KM Human placenta; metalloproteinase; probe; cancer cell detection;
 KM monoclonal antibody; cancer; treatment; diagnosis.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 112..1860

FT W09515374-A1. /tag- a
 PN 08-JUN-1995.
 PD 30-NOV-1994; J02009.
 PR 30-NOV-1993; JP-341061.
 PA (FUJII) FUJII YAKUHI KOGYO KK.
 PI Sato H, Seki M, Shinagawa A;
 DR WPI: 95-215253/28.
 N-PSDB: Q92573.
 PT Metalloprotease and monoclonal antibody recognising it - for
 detection of cancer cells in medical diagnosis and research.
 PS Claim 1: Pages 39-43; 67pp; Japanese.
 CC 092573 encodes R75648 a human placenta derived metalloprotease. The
 gene can be used as a probe for the detection of cancer cells, and
 CC monoclonal antibodies specific for the metalloprotease can be used
 for the treatment and diagnosis of cancer.
 SO Sequence 582 AA;

Query Match 45.28; Score 61; DB 14; Length 582;
 Best Local Similarity 29.48; Pred. No. 2,77e+01;
 Matches 5; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

DB 163 pyayireghekqadlm 179
 :||||:|:|:|:
 QY 3 SYSFIRTAHSEVQDLIA 19

RESULT 9
 ID R32010 standard; Protein: 146 AA.
 AC R32010:
 DT 27-MAY-1993 (first entry)
 DE Rpl3-T1A-1.
 KW 15 kD antigen; cytolytic lymphocyte; infection; HIV; targeting.
 OS Homo sapiens.
 PN W09301314-A.
 PD 21-JAN-1993.
 PF 12-JUN-1992; U05117.
 PR 10-JUL-1991; US-726607.
 PR 19-FEB-1992; US-843949.
 PA (DAND) DANA FARBER CANCER INST INC.
 PI Anderson PJ, Schlossman SF, Streuli M;
 DR WPI: 93-045515/05.
 N-PSDB: Q34555.
 PT DNA encoding T1A-1 antigen - used for detecting and cytotoxic
 PT killing of cytolytic lymphocyte(s) applicable for cancer
 PT treatment
 PS Disclosure: Page 37; 61pp; English.
 CC RNA from cytolytic T cell clone T4R8C1, which expresses high levels
 of T1A-1 antigen was used for the construction of a cDNA library in
 CC lambda gtl1. When this expression library was screened using T1A-1
 CC monoclonal antibody, several bacteriophage expressing immunoreactive
 CC fusion proteins were identified. Cross hybridisation analysis
 CC identified three independent clones contg. related cDNAs. The
 CC largest cDNA insert was used to probe the original library for
 CC isolation of full length cDNA encoding rpl3-T1A-1, a 15 kD T1A-1
 CC antigen. The antigen is active in lymphocyte-mediated cytotoxicity
 CC and can eliminate a wide variety of virus infected or transformed
 CC target cells. T1A-1 or T1A-1 antigen may be linked chemically or
 CC concomitantly to cell targeting ligands such as growth factors,
 CC hormone or antibodies, and may be used to kill targeted cells, e.g.
 CC cancer cells. See also R32009-17.
 SO Sequence 146 AA;

Query Match 44.48; Score 60; DB 6; Length 146;
 Best Local Similarity 31.68; Pred. No. 3.53e+01;
 Matches 6; Conservative 8; Mismatches 4; Indels 1; Gaps 1;

DB 10 gysfvrfnshesaahivs 28
 :|||:|:|:|:|:
 QY 3 SYSFIR-TAHSEVQDLIA 20

RESULT 10

ID R12539 standard; Protein: 146 AA.
 AC R12539;
 DT 24-SEP-1991 (first entry)
 DE Cytotoxic T lymphocyte-specific antigen.
 KW CTL: natural killer cells; CD8; peripheral blood; T1A-1; immunogen.
 PN EP-436400-A.
 PD 10-JUL-1991.
 PF 31-DEC-1990; 314456.
 PR 05-JAN-1990; US-460678.
 PA (DANA-) DANA FARBER CANCER.
 PI Anderson PJ, Streuli M, Schlossman SF;
 DR WPI: 91-202096/28.
 N-PSDB: Q12525.
 PT Monoclonal antibody to identify cytolytic lymphocytes - reactive
 PT with a 15kD protein in cytolytic T lymphocytes and natural killer
 PT cells
 PS Claim 10: Fig 1; 11pp; English.
 CC This sequence corresponds to a 15kD immunogenic protein associated
 CC with cytoplasmic granules in cytolytic T lymphocytes and natural
 CC killer cells. The protein is found principally in a subpopulation of
 CC CD8+ T lymphocytes from peripheral blood mononuclear cells. It
 CC reacts with monoclonal antibody T1A-1 (HB 10319).
 SO Sequence 146 AA;

Query Match 44.48; Score 60; DB 3; Length 146;
 Best Local Similarity 31.68; Pred. No. 3.53e+01;
 Matches 6; Conservative 8; Mismatches 4; Indels 1; Gaps 1;

DB 10 gysfvrfnshesaahivs 28
 :|||:|:|:|:|:
 QY 3 SYSFIR-TAHSEVQDLIA 20

RESULT 11
 ID R40840 standard; Protein: 156 AA.
 AC R40840;
 DT 01-MAR-1994 (first entry)
 DE Glioblastoma derived polypeptide.
 KW Glioblastoma; differentiation; proliferation; growth; glias;
 KW neuron; immunity; tumour; aplasia.
 OS Homo sapiens.
 FH Key
 FT peptide
 FT 1..23
 FT /label= Signal_peptide
 FT 24..156
 FT /label= Mature_protein
 PN EP-559428-A.
 PD 08-SEP-1993.
 PF 02-MAR-1993; 301571.
 PR 03-MAR-1992; JP-081557.
 PR 22-APR-1992; JP-129558.
 PA (ONOY) ONO PHARM CO LTD.
 PI Konishi M, Miyamoto T, Naitoh T;
 DR WPI: 93-282254/36.
 N-PSDB: Q47756.
 PT New polypeptide from glioblastoma cell line - useful in treatment
 PT of abnormal glial cells or neurons or for treating tumours
 PS Claims 1-4: Page 14-15; 21pp; English.
 CC The glioblastoma derived polypeptide is synthesised in and secreted
 CC from a glioblastoma cell line containing this DNA fragment and
 CC therefore may possess biological activities relating to
 CC differentiation, proliferation and growth of glias or neurons,
 CC relating to the function of immunity and relating to proliferation
 CC and growth of tumours. The polypeptide may be useful for the
 CC prevention of or in the treatment of aplasia or abnormal
 CC proliferation of glias or neurons, depression or enhancement of
 CC immunological activity. The DNA was amplified by PCR techniques
 CC using primers (Q47754-55).
 SO Sequence 156 AA;

Query Match 44.48; Score 60; DB 8; Length 156;
 Best Local Similarity 42.98; Pred. No. 3.53e+01;
 Matches 9; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

PS Disclosure; Page 35; 61pp; English

KM Catalysis; enzyme; methyl group; vanillic acid; syringa; production.
KM catechol; pyrogarol; *Pseudomonas paucimobilis*.
OS *Pseudomonas paucimobilis* SYK-6.

PN J09173074-A.
 PD 08-JUL-1997.
 PF 21-DEC-1995; 349914.
 PR 21-DEC-1995; JP-349914.
 PA (MAZN) COSMO OIL CO LTD.
 PA (COSM-) COSMO SOGO KENKYUSHO KK.
 DR MPI; 97-397032/37.
 DR N-PSDB; T85645.
 PT Gene encoding enzyme which catalyses release of methyl group from
 PT vanillic or syringic acid - useful for production of catechol or
 PT pyrogallol containing molecules
 PS Claim 2: Page 8-9: 14pp: Japanese.
 CC This is an enzyme which is capable of catalysing the reaction of
 CC releasing a methyl group from vanillic acid or syringic acid. The DNA and
 CC transformed host cells can be used to produce a recombinant enzyme. This
 CC enzyme plays an important role in transforming natural components into
 CC those of catechol or pyrogallol structure.
 SO Sequence 557 AA;

Very Match 43.78; Score 59; DB 24; Length 557;
 Best Local Similarity 36.88; Pred. No. 4.48e+01;
 Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Db 539 vpsadffikldggqiglf 557
 ||| ||: :||: |
 Oy 1 VPSYSFIRTAHDSVODLI 19

Search completed: Tue Jan 11 15:41:38 2000
 Job time : 21 secs.

WIDEORLH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Jan 11 15:46:03 2000; MasPar time 1.64 Seconds
145,493 Million cell updates/sec
Similar output not generated.

Description: >US-09-290-049-2
(1-20) from US09290049.pep
Perfect Score: 135
Sequence: 1 VPSYSPFRTAHSEVODLIA 20

Scoring table: PAM 150
Gap 15

Searched: 122461 seqs, 11912985 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A.COMB 2:5B.COMB 3:PCT9.COMB 4:backfiles1

Statistics: Mean 19.113; Variance 59.268; scale 0.322

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	61	45.2	344	2	US-08-689-Sequence 1, Applicatio	1.38e+01
2	57	42.2	160	2	US-08-726-Sequence 183, Applicat	3.61e+01
3	57	42.2	194	2	US-08-689-Sequence 3, Applicatio	3.61e+01
4	56	41.5	366	2	US-08-690-Sequence 96, Applicatio	4.58e+01
5	55	40.7	256	2	US-08-469-Sequence 58, Applicati	5.80e+01
6	55	40.7	266	2	US-07-857-Sequence 75, Applicatio	5.80e+01
7	55	40.7	282	2	US-08-701-Sequence 19, Applicatio	5.80e+01
8	55	40.7	1003	2	US-07-743-Sequence 9, Applicatio	7.33e+01
9	54	40.0	244	2	US-08-553-Sequence 2, Applicatio	7.33e+01
10	54	40.0	244	2	US-08-737-Sequence 2, Applicatio	7.33e+01
11	54	40.0	473	2	US-08-746-Sequence 13, Applicati	7.33e+01
12	54	40.0	473	1	US-08-597-Sequence 13, Applicati	7.33e+01
13	54	40.0	1069	2	US-08-162-Sequence 37, Applicati	7.33e+01
14	54	40.0	1069	2	US-08-780-Sequence 37, Applicati	7.33e+01
15	54	40.0	1080	2	US-08-780-Sequence 36, Applicati	7.33e+01
16	54	40.0	1080	2	US-08-162-Sequence 36, Applicati	7.33e+01
17	53	39.3	980	1	US-08-413-Sequence 5, Applicatio	9.24e+01
18	53	39.3	980	1	US-08-220-Sequence 5, Applicatio	9.24e+01
19	53	39.3	1172	2	US-08-313-Sequence 19, Applicati	9.24e+01
20	52	38.5	418	3	PCT-US94-0-Sequence 72, Applicati	1.16e+02
21	52	38.5	865	1	US-07-803-Sequence 13, Applicati	1.16e+02
22	52	38.5	913	2	US-07-743-Sequence 22, Applicati	1.16e+02
23	52	38.5	913	2	US-07-743-Sequence 6, Applicatio	1.16e+02

24	52	38.5	1003	2	US-07-743-Sequence 10, Applicati	1.16e+02
25 <td>52<td>38.5<td>1003<td>2<td>US-07-743-Sequence 8, Applicatio<th>1.16e+02</th></td></td></td></td></td>	52 <td>38.5<td>1003<td>2<td>US-07-743-Sequence 8, Applicatio<th>1.16e+02</th></td></td></td></td>	38.5 <td>1003<td>2<td>US-07-743-Sequence 8, Applicatio<th>1.16e+02</th></td></td></td>	1003 <td>2<td>US-07-743-Sequence 8, Applicatio<th>1.16e+02</th></td></td>	2 <td>US-07-743-Sequence 8, Applicatio<th>1.16e+02</th></td>	US-07-743-Sequence 8, Applicatio <th>1.16e+02</th>	1.16e+02
26 <td>52<td>38.5<td>1004<td>2<td>US-07-743-Sequence 7, Applicatio<th>1.16e+02</th></td></td></td></td></td>	52 <td>38.5<td>1004<td>2<td>US-07-743-Sequence 7, Applicatio<th>1.16e+02</th></td></td></td></td>	38.5 <td>1004<td>2<td>US-07-743-Sequence 7, Applicatio<th>1.16e+02</th></td></td></td>	1004 <td>2<td>US-07-743-Sequence 7, Applicatio<th>1.16e+02</th></td></td>	2 <td>US-07-743-Sequence 7, Applicatio<th>1.16e+02</th></td>	US-07-743-Sequence 7, Applicatio <th>1.16e+02</th>	1.16e+02
27 <td>52<td>38.5<td>1005<td>2<td>US-07-743-Sequence 1, Applicatio<th>1.16e+02</th></td></td></td></td></td>	52 <td>38.5<td>1005<td>2<td>US-07-743-Sequence 1, Applicatio<th>1.16e+02</th></td></td></td></td>	38.5 <td>1005<td>2<td>US-07-743-Sequence 1, Applicatio<th>1.16e+02</th></td></td></td>	1005 <td>2<td>US-07-743-Sequence 1, Applicatio<th>1.16e+02</th></td></td>	2 <td>US-07-743-Sequence 1, Applicatio<th>1.16e+02</th></td>	US-07-743-Sequence 1, Applicatio <th>1.16e+02</th>	1.16e+02
28 <td>52<td>38.5<td>1016<td>2<td>US-07-743-Sequence 2, Applicatio<th>1.16e+02</th></td></td></td></td></td>	52 <td>38.5<td>1016<td>2<td>US-07-743-Sequence 2, Applicatio<th>1.16e+02</th></td></td></td></td>	38.5 <td>1016<td>2<td>US-07-743-Sequence 2, Applicatio<th>1.16e+02</th></td></td></td>	1016 <td>2<td>US-07-743-Sequence 2, Applicatio<th>1.16e+02</th></td></td>	2 <td>US-07-743-Sequence 2, Applicatio<th>1.16e+02</th></td>	US-07-743-Sequence 2, Applicatio <th>1.16e+02</th>	1.16e+02
29 <td>52<td>38.5<td>1016<td>2<td>US-07-743-Sequence 3, Applicatio<th>1.16e+02</th></td></td></td></td></td>	52 <td>38.5<td>1016<td>2<td>US-07-743-Sequence 3, Applicatio<th>1.16e+02</th></td></td></td></td>	38.5 <td>1016<td>2<td>US-07-743-Sequence 3, Applicatio<th>1.16e+02</th></td></td></td>	1016 <td>2<td>US-07-743-Sequence 3, Applicatio<th>1.16e+02</th></td></td>	2 <td>US-07-743-Sequence 3, Applicatio<th>1.16e+02</th></td>	US-07-743-Sequence 3, Applicatio <th>1.16e+02</th>	1.16e+02
30 <td>52<td>38.5<td>1016<td>2<td>US-07-743-Sequence 4, Applicatio<th>1.16e+02</th></td></td></td></td></td>	52 <td>38.5<td>1016<td>2<td>US-07-743-Sequence 4, Applicatio<th>1.16e+02</th></td></td></td></td>	38.5 <td>1016<td>2<td>US-07-743-Sequence 4, Applicatio<th>1.16e+02</th></td></td></td>	1016 <td>2<td>US-07-743-Sequence 4, Applicatio<th>1.16e+02</th></td></td>	2 <td>US-07-743-Sequence 4, Applicatio<th>1.16e+02</th></td>	US-07-743-Sequence 4, Applicatio <th>1.16e+02</th>	1.16e+02
31 <td>51<td>37.8<td>221<td>2<td>US-08-916-Sequence 9, Applicatio<th>1.46e+02</th></td></td></td></td></td>	51 <td>37.8<td>221<td>2<td>US-08-916-Sequence 9, Applicatio<th>1.46e+02</th></td></td></td></td>	37.8 <td>221<td>2<td>US-08-916-Sequence 9, Applicatio<th>1.46e+02</th></td></td></td>	221 <td>2<td>US-08-916-Sequence 9, Applicatio<th>1.46e+02</th></td></td>	2 <td>US-08-916-Sequence 9, Applicatio<th>1.46e+02</th></td>	US-08-916-Sequence 9, Applicatio <th>1.46e+02</th>	1.46e+02
32 <td>51<td>37.8<td>221<td>2<td>US-08-773-Sequence 9, Applicatio<th>1.46e+02</th></td></td></td></td></td>	51 <td>37.8<td>221<td>2<td>US-08-773-Sequence 9, Applicatio<th>1.46e+02</th></td></td></td></td>	37.8 <td>221<td>2<td>US-08-773-Sequence 9, Applicatio<th>1.46e+02</th></td></td></td>	221 <td>2<td>US-08-773-Sequence 9, Applicatio<th>1.46e+02</th></td></td>	2 <td>US-08-773-Sequence 9, Applicatio<th>1.46e+02</th></td>	US-08-773-Sequence 9, Applicatio <th>1.46e+02</th>	1.46e+02
33 <td>51<td>37.8<td>375<td>1<td>US-08-027-Sequence 2, Applicatio<th>1.46e+02</th></td></td></td></td></td>	51 <td>37.8<td>375<td>1<td>US-08-027-Sequence 2, Applicatio<th>1.46e+02</th></td></td></td></td>	37.8 <td>375<td>1<td>US-08-027-Sequence 2, Applicatio<th>1.46e+02</th></td></td></td>	375 <td>1<td>US-08-027-Sequence 2, Applicatio<th>1.46e+02</th></td></td>	1 <td>US-08-027-Sequence 2, Applicatio<th>1.46e+02</th></td>	US-08-027-Sequence 2, Applicatio <th>1.46e+02</th>	1.46e+02
34 <td>51<td>37.8<td>375<td>1<td>US-08-827-Sequence 1, Applicatio<th>1.46e+02</th></td></td></td></td></td>	51 <td>37.8<td>375<td>1<td>US-08-827-Sequence 1, Applicatio<th>1.46e+02</th></td></td></td></td>	37.8 <td>375<td>1<td>US-08-827-Sequence 1, Applicatio<th>1.46e+02</th></td></td></td>	375 <td>1<td>US-08-827-Sequence 1, Applicatio<th>1.46e+02</th></td></td>	1 <td>US-08-827-Sequence 1, Applicatio<th>1.46e+02</th></td>	US-08-827-Sequence 1, Applicatio <th>1.46e+02</th>	1.46e+02
35 <td>51<td>37.8<td>393<td>2<td>US-08-689-Sequence 4, Applicatio<th>1.46e+02</th></td></td></td></td></td>	51 <td>37.8<td>393<td>2<td>US-08-689-Sequence 4, Applicatio<th>1.46e+02</th></td></td></td></td>	37.8 <td>393<td>2<td>US-08-689-Sequence 4, Applicatio<th>1.46e+02</th></td></td></td>	393 <td>2<td>US-08-689-Sequence 4, Applicatio<th>1.46e+02</th></td></td>	2 <td>US-08-689-Sequence 4, Applicatio<th>1.46e+02</th></td>	US-08-689-Sequence 4, Applicatio <th>1.46e+02</th>	1.46e+02
36 <td>51<td>37.8<td>445<td>2<td>US-08-630-Sequence 6, Applicatio<th>1.46e+02</th></td></td></td></td></td>	51 <td>37.8<td>445<td>2<td>US-08-630-Sequence 6, Applicatio<th>1.46e+02</th></td></td></td></td>	37.8 <td>445<td>2<td>US-08-630-Sequence 6, Applicatio<th>1.46e+02</th></td></td></td>	445 <td>2<td>US-08-630-Sequence 6, Applicatio<th>1.46e+02</th></td></td>	2 <td>US-08-630-Sequence 6, Applicatio<th>1.46e+02</th></td>	US-08-630-Sequence 6, Applicatio <th>1.46e+02</th>	1.46e+02
37 <td>51<td>37.8<td>445<td>2<td>US-08-838-Sequence 6, Applicatio<th>1.46e+02</th></td></td></td></td></td>	51 <td>37.8<td>445<td>2<td>US-08-838-Sequence 6, Applicatio<th>1.46e+02</th></td></td></td></td>	37.8 <td>445<td>2<td>US-08-838-Sequence 6, Applicatio<th>1.46e+02</th></td></td></td>	445 <td>2<td>US-08-838-Sequence 6, Applicatio<th>1.46e+02</th></td></td>	2 <td>US-08-838-Sequence 6, Applicatio<th>1.46e+02</th></td>	US-08-838-Sequence 6, Applicatio <th>1.46e+02</th>	1.46e+02
38 <td>51<td>37.8<td>455<td>1<td>US-08-349-Sequence 4, Applicatio<th>1.46e+02</th></td></td></td></td></td>	51 <td>37.8<td>455<td>1<td>US-08-349-Sequence 4, Applicatio<th>1.46e+02</th></td></td></td></td>	37.8 <td>455<td>1<td>US-08-349-Sequence 4, Applicatio<th>1.46e+02</th></td></td></td>	455 <td>1<td>US-08-349-Sequence 4, Applicatio<th>1.46e+02</th></td></td>	1 <td>US-08-349-Sequence 4, Applicatio<th>1.46e+02</th></td>	US-08-349-Sequence 4, Applicatio <th>1.46e+02</th>	1.46e+02
39 <td>51<td>37.8<td>456<td>3<td>PCT-US95-1-Sequence 4, Applicatio<th>1.46e+02</th></td></td></td></td></td>	51 <td>37.8<td>456<td>3<td>PCT-US95-1-Sequence 4, Applicatio<th>1.46e+02</th></td></td></td></td>	37.8 <td>456<td>3<td>PCT-US95-1-Sequence 4, Applicatio<th>1.46e+02</th></td></td></td>	456 <td>3<td>PCT-US95-1-Sequence 4, Applicatio<th>1.46e+02</th></td></td>	3 <td>PCT-US95-1-Sequence 4, Applicatio<th>1.46e+02</th></td>	PCT-US95-1-Sequence 4, Applicatio <th>1.46e+02</th>	1.46e+02
40 <td>51<td>37.8<td>598<td>2<td>US-08-837-Sequence 2, Applicatio<th>1.46e+02</th></td></td></td></td></td>	51 <td>37.8<td>598<td>2<td>US-08-837-Sequence 2, Applicatio<th>1.46e+02</th></td></td></td></td>	37.8 <td>598<td>2<td>US-08-837-Sequence 2, Applicatio<th>1.46e+02</th></td></td></td>	598 <td>2<td>US-08-837-Sequence 2, Applicatio<th>1.46e+02</th></td></td>	2 <td>US-08-837-Sequence 2, Applicatio<th>1.46e+02</th></td>	US-08-837-Sequence 2, Applicatio <th>1.46e+02</th>	1.46e+02
41 <td>51<td>37.8<td>718<td>1<td>US-08-445-Sequence 4, Applicatio<th>1.46e+02</th></td></td></td></td></td>	51 <td>37.8<td>718<td>1<td>US-08-445-Sequence 4, Applicatio<th>1.46e+02</th></td></td></td></td>	37.8 <td>718<td>1<td>US-08-445-Sequence 4, Applicatio<th>1.46e+02</th></td></td></td>	718 <td>1<td>US-08-445-Sequence 4, Applicatio<th>1.46e+02</th></td></td>	1 <td>US-08-445-Sequence 4, Applicatio<th>1.46e+02</th></td>	US-08-445-Sequence 4, Applicatio <th>1.46e+02</th>	1.46e+02
42 <td>51<td>37.8<td>788<td>2<td>US-07-728-Sequence 32, Applicati<th>1.46e+02</th></td></td></td></td></td>	51 <td>37.8<td>788<td>2<td>US-07-728-Sequence 32, Applicati<th>1.46e+02</th></td></td></td></td>	37.8 <td>788<td>2<td>US-07-728-Sequence 32, Applicati<th>1.46e+02</th></td></td></td>	788 <td>2<td>US-07-728-Sequence 32, Applicati<th>1.46e+02</th></td></td>	2 <td>US-07-728-Sequence 32, Applicati<th>1.46e+02</th></td>	US-07-728-Sequence 32, Applicati <th>1.46e+02</th>	1.46e+02
43 <td>51<td>37.8<td>1082<td>1<td>US-08-106-Sequence 2, Applicatio<th>1.46e+02</th></td></td></td></td></td>	51 <td>37.8<td>1082<td>1<td>US-08-106-Sequence 2, Applicatio<th>1.46e+02</th></td></td></td></td>	37.8 <td>1082<td>1<td>US-08-106-Sequence 2, Applicatio<th>1.46e+02</th></td></td></td>	1082 <td>1<td>US-08-106-Sequence 2, Applicatio<th>1.46e+02</th></td></td>	1 <td>US-08-106-Sequence 2, Applicatio<th>1.46e+02</th></td>	US-08-106-Sequence 2, Applicatio <th>1.46e+02</th>	1.46e+02
44 <td>51<td>37.8<td>1139<td>2<td>US-08-832-Sequence 2, Applicatio<th>1.46e+02</th></td></td></td></td></td>	51 <td>37.8<td>1139<td>2<td>US-08-832-Sequence 2, Applicatio<th>1.46e+02</th></td></td></td></td>	37.8 <td>1139<td>2<td>US-08-832-Sequence 2, Applicatio<th>1.46e+02</th></td></td></td>	1139 <td>2<td>US-08-832-Sequence 2, Applicatio<th>1.46e+02</th></td></td>	2 <td>US-08-832-Sequence 2, Applicatio<th>1.46e+02</th></td>	US-08-832-Sequence 2, Applicatio <th>1.46e+02</th>	1.46e+02
45 <td>51<td>37.8<td>1139<td>2<td>US-08-832-Sequence 2, Applicatio<th>1.46e+02</th></td></td></td></td></td>	51 <td>37.8<td>1139<td>2<td>US-08-832-Sequence 2, Applicatio<th>1.46e+02</th></td></td></td></td>	37.8 <td>1139<td>2<td>US-08-832-Sequence 2, Applicatio<th>1.46e+02</th></td></td></td>	1139 <td>2<td>US-08-832-Sequence 2, Applicatio<th>1.46e+02</th></td></td>	2 <td>US-08-832-Sequence 2, Applicatio<th>1.46e+02</th></td>	US-08-832-Sequence 2, Applicatio <th>1.46e+02</th>	1.46e+02

ALIGNMENTS

RESULT 1
ID US-08-689-974-1 STANDARD; PRT: 344 AA.

Sequence 1, Application US/08689974

CC Patent No. 5776732
CC Sequence 1, Application US/08689974

CC GENERAL INFORMATION:

CC APPLICANT: Au-Young, Janice

CC APPLICANT: Hawkins, Phillip R.

CC APPLICANT: Murray, Lynn E.

CC TITLE OF INVENTION: NOVEL HUMAN INDUCED TUMOR PROTEIN

CC NUMBER OF SEQUENCES: 5

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Incyte Pharmaceuticals, Inc.

CC STREET: 3174 Porter Drive

CC CITY: Palo Alto

CC STATE: CA

CC COUNTRY: U.S.

CC ZIP: 94304

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Diskette

CC OPERATING SYSTEM: DOS

CC SOFTWARE: FASTSEQ Version 1.5

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/689,974

CC FILING DATE: Filed Herewith

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Billings, Lucy J.

CC REGISTRATION NUMBER: 36,749

CC REFERENCE/DOCKET NUMBER: PF-0113 US

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: 415-855-0555

CC TELEFAX: 415-845-4166

CC INFORMATION FOR SEQ ID NO: 1:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 344 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

CC IMMEDIATE SOURCE:


```

Db      98 SGYTFVLHHDHNNVATL 114
      :|:|:|:|:|
QY      2 PSYSEFIRAHDSQVDL 18

RESULT 10
ID      US-08-737-825-2      STANDARD:      PRT:      244 AA.
XX      AC      xxxxxx
XX
XX
XX      Sequence 2, Application US/08737825
DE
CC      Sequence 2, Application US/08737825
CC      Patent No. 5871922
CC      GENERAL INFORMATION:
CC      APPLICANT: SALMOND, GEORGE PEACOCK COPELAND
CC      APPLICANT: MCGOMAN, SIMON JAMES
CC      APPLICANT: SEBAIHA, MOHAMED
CC      APPLICANT: COX, ANTHONY RICHARD JOHN
CC      APPLICANT: HOLDEN, MATTHEW THOMAS GEOFFREY
CC      APPLICANT: PORTER, LAUREN ELIZABETH
CC      APPLICANT: BYCROFT, BARRIE WALSHAM
CC      APPLICANT: WILLIAMS, PAUL
CC      APPLICANT: STEWART, GORDON SIDNEY ANDERSON BIRNIE
CC      TITLE OF INVENTION: GENES INVOLVED IN THE BIOSYNTHETIC PATHWAY OF CARBAPENEM
CC      NUMBER OF SEQUENCES: 12
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
CC      STREET: P.O. Box 747
CC      CITY: Falls Church
CC      STATE: Virginia
CC      COUNTRY: USA
CC      ZIP: 22040-0747
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: Floppy disk
CC      COMPUTER: IBM PC compatible
CC      OPERATING SYSTEM: PC-DOS/MS-DOS
CC      SOFTWARE: MICROSOFT WORD97
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: US/08/737,825
CC      FILING DATE: 03-JUN-1997
CC      CLASSIFICATION: 435
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: SVENSSON, LEONARD R.
CC      REGISTRATION NUMBER: 30,330
CC      REFERENCE/DOCKET NUMBER: 1009-0105P
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: (703) 205-8000
CC      TELEFAX: (703) 205-8050
CC      TELEX: 248345
CC      INFORMATION FOR SEQ ID NO: 2:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 244 amino acids
CC      TYPE: amino acid
CC      STRANDEDNESS:
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: protein
CC      HYPOTHEetical: NO
CC      ANTI-SENSE: NO
CC      ORIGINAL SOURCE:
CC      ORGANISM: Erwinia carotovora
CC      SEQUENCE 244 AA; 28137 MW; 305391 CN;
SQ

Query Match      40.0%; Score 54; DB 2; Length 244;
Best Local Similarity 35.3%; Pred. No. 7.33e+01;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;
      98 SGYTFVLHHDHNNVATL 114

```

OY 2 PSYSFIRTAHSEVODL 18

RESULT 11
ID US-08-746-682A-13 STANDARD: PRT: 473 AA.

Sequence 13, Application US/08746682A

Sequence 13, Application US/08746682A
Patent No. 5786184

GENERAL INFORMATION:

APPLICANT: STINGELE, Francesca

APPLICANT: MOLIET, Beat

TITLE OF INVENTION: LACTIC BACTERIA PRODUCING

TITLE OF INVENTION: EXOPOLYSACCHARIDES

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americans

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/746,682A

FILING DATE: 14-NOV-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/597,236

FILING DATE: 20-JUN-1995

APPLICATION NUMBER: EP 95201669.9

FILING DATE: 20-JUN-1995

ATTORNEY/AGENT INFORMATION:

NAME: Fanucci A., Allan

REGISTRATION NUMBER: 30256

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 473 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 473 AA; 53028 MW; 1298864 CN;

Query Match 40.0%; Score 54; DB 2; Length 473;

Best Local Similarity 33.3%; Pred. No. 7.33e+01;

Matches 6; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

Db 29 VPLYTAMTPOEYGMADL 46

1 VPSYSFIRTAHSEVODL 18

RESULT 12
ID US-08-597-236-13 STANDARD: PRT: 473 AA.

Sequence 13, Application US/08597236

Sequence 13, Application US/08597236

Patent No. 5733765

Patent No. 5733765

Patent No. 5733765

Patent No. 5733765

Patent No. 5733765

Patent No. 5733765

Patent No. 5733765

Patent No. 5733765

GENERAL INFORMATION:

APPLICANT: STINGELE, Francesca

APPLICANT: MOLIET, Beat

TITLE OF INVENTION: LACTIC BACTERIA PRODUCING

TITLE OF INVENTION: EXOPOLYSACCHARIDES

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americans

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/597,236

FILING DATE:

CLASSIFICATION: 426

PRIOR APPLICATION DATA:

APPLICATION NUMBER: EP 95201669.9

FILING DATE: 20-JUN-1995

ATTORNEY/AGENT INFORMATION:

NAME: Fanucci A., Allan

REGISTRATION NUMBER: 30256

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 473 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 473 AA; 53028 MW; 1298864 CN;

Query Match 40.0%; Score 54; DB 1; Length 473;

Best Local Similarity 33.3%; Pred. No. 7.33e+01;

Matches 6; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Db 29 VPLYTAMTPOEYGMADL 46

1 VPSYSFIRTAHSEVODL 18

RESULT 13
ID US-08-162-081B-37 STANDARD: PRT: 1069 AA.

Sequence 37, Application US/08162081B

Sequence 37, Application US/08162081B

Patent No. 5824492

GENERAL INFORMATION:

APPLICANT: Hilles, Ian Donald; Fry, Michael John; Dhand, Ritu

APPLICANT: Balas, Waterfield, Michael Derek; Parker, Peter

APPLICANT: Joseph, Otsu, Masayuki; Panayotou, George; Volintia,

APPLICANT: Stefano, Gout, Ivan Tarasovitch

TITLE OF INVENTION: POLYPEPTIDES HAVING KINASE ACTIVITY,

TITLE OF INVENTION: THEIR PREPARATION AND USE

NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:

ADDRESSEE: Felte & Lynch

STREET: 805 Third Avenue

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage

```
CC COMPUTER: IBM PS/2
CC OPERATING SYSTEM: PC-DOS
CC SOFTWARE: Wordperfect
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/162,081B
CC FILING DATE: February 7, 1994
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: PCT/GB93/00761
CC FILING DATE: 13 April 1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Pasqualini, Patricia A.
CC REGISTRATION NUMBER: 34,894
CC REFERENCE/DOCKET NUMBER: LUD 5256
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 688-9200
CC TELEFAX: (212) 838-3884
CC INFORMATION FOR SEO ID NO: 37:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1069 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC SEQUENCE 1069 AA; 124437 MW; 5707845 CN;

Query Match 40.0%; Score 54; DB 2; Length 1069;
Best Local Similarity 22.2%; Pred. No. 7.33e+01;
Matches 4; Conservative 10; Mismatches 4; Indels 0; Gaps 0;

Db 66 SSYIFVSYTOAEAREEFF 83
QY 2 PSYSFIPTAHDSVEVDLI 19

RESULT 14
ID US-08-780-872-37 STANDARD: PRT; 1069 AA.
XX
AC xxxxxx
AC
DT
XX
XX
XX
DE Sequence 37, Application US/08780872

Sequence 37, Application US/08780872
Patent No. 5846824
GENERAL INFORMATION:
APPLICANT: Hiles, Ian Donald; Fry, Michael John; Dhand, Ritu
APPLICANT: Bala; Waterfield, Michael Derek; Parker, Peter
APPLICANT: Joseph, Otsu, Masayuki; Panayotou, George; Volina,
TITLE OF INVENTION: Stefano; Gout, Ivan Tarasovitch
TITLE OF INVENTION: POLYPEPTIDES HAVING KINASE ACTIVITY,
NUMBER OF INVENTIONS: THEIR PREPARATION AND USE
CORRESPONDENCE ADDRESSES: 50
ADDRESSEE: felle & Lynch
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/780,872
FILING DATE: 09-JAN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/162,081
FILING DATE: February 7, 1994
APPLICATION NUMBER: PCT/GB93/00761
```

CC	FILING DATE:	13 April 1993
CC	ATTORNEY/AGENT INFORMATION:	
CC	NAME:	Pasqualini, Patricia A.
CC	REGISTRATION NUMBER:	34,894
CC	REFERENCE/DOCKET NUMBER:	LUD 5256
CC	TELECOMMUNICATION INFORMATION:	
CC	TELEPHONE:	(212) 688-9200
CC	TELEFAX:	(212) 838-3884
CC	INFORMATION FOR SEQ ID NO:	37:
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH:	1069 amino acids
CC	TYPE:	amino acid
CC	STRANDEDNESS:	single
CC	TOPOLOGY:	linear
SQ	SEQUENCE	1069 AA; 124437 MW; 5707845 CN;
Db	Query Match	40.0%; Score 54; DB 2; Length 1069;
Oy	Best Local Similarity	22.2%; Pred. No. 7.33e+01;
	Matches	4; Conservative 10; Mismatches 4; Indels 0; Gaps 0;
Db	66 SSIIIVSYTQAEAREEEFF 83	
	:: ::: :::	
Oy	2 PSYSIFRTAHDSVEVDLTI 19	
RESULT	15	
ID	US-08-780-872-36	STANDARD; PRT; 1080 AA.
AC	xxxxxx	
XX		
XX		
DE	Sequence 36, Application US/08780872	
CC	Sequence 36, Application US/08780872	
CC	Patent No. 5846824	
CC	GENERAL INFORMATION:	
CC	APPLICANT:	Hiles, Ian Donald; Fry, Michael John; Dhand, Ritu
CC	APPLICANT:	Bala; Waterfield, Michael Derek; Parker, Peter
CC	APPLICANT:	Joseph; Otsu, Masayuki; Panayotou, George; Volinaia,
CC	APPLICANT:	Stefano; Gout, Ivan Tarasovich
CC	TITLE OF INVENTION:	POLYPEPTIDES HAVING KINASE ACTIVITY,
CC	TITLE OF INVENTION:	THEIR PREPARATION AND USE
CC	NUMBER OF SEQUENCES:	50
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESSEE:	Felfe & Lynch
CC	STREET:	805 Third Avenue
CC	CITY:	New York
CC	STATE:	New York
CC	COUNTRY:	USA
CC	ZIP:	10022
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE:	Diskette, 5.25 inch, 360 kb storage
CC	COMPUTER:	IBM PS/2
CC	OPERATING SYSTEM:	PC-DOS
CC	SOFTWARE:	Wordperfect
CC	CURRENT APPLICATION DATA:	
CC	APPLICATION NUMBER:	US/08/780,872
CC	FILING DATE:	09-JAN-1997
CC	CLASSIFICATION:	435
CC	PRIOR APPLICATION DATA:	
CC	APPLICATION NUMBER:	08/162,081
CC	FILING DATE:	February 7, 1994
CC	APPLICATION NUMBER:	PCT/GB93/00761
CC	FILING DATE:	13 April 1993
CC	ATTORNEY/AGENT INFORMATION:	
CC	NAME:	Pasqualini, Patricia A.
CC	REGISTRATION NUMBER:	34,894
CC	REFERENCE/DOCKET NUMBER:	LUD 5256
CC	TELECOMMUNICATION INFORMATION:	
CC	TELEPHONE:	(212) 688-9200
CC	TELEFAX:	(212) 838-3884
CC	INFORMATION FOR SEQ ID NO:	36

CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 1080 amino acids
 CC TYPE: amino acid
 CC STRANDEDNESS: single
 CC TOPOLOGY: linear
 SQ SEQUENCE 1080 AA; 125733 MW; 5831251 CN;

Query Match 40.0%; Score 54; DB 2; Length 1080;
 Best Local Similarity 22.2%; Pred. No. 7.33e+01;
 Matches 4; Conservative 10; Mismatches 4; Indels 0; Gaps 0;
 Db 66 SSYIFVSVTOAEREFE 83
 :|||: :|||: :|||: :|||: :|||:
 QY 2 PSISFIRAHDSVQDLI 19

Search completed: Tue Jan 11 15:46:11 2000
 time : 8 secs.

 WISE (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit,
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Merch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Tue Jan 11 15:41:54 2000; Maspar time 3.73 Seconds
 ar output not generated. 214,842 Million cell updates/sec

Title: >US-09-290-049-2
 Description: (1-20) from US09290049.pep
 Perfect Score: 135
 Sequence: 1 VPYSFIRTAHDSFVODLIA 20

Scoring table: PAM 150
 Gap 15

Searched: 122810 segs, 40068593 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: p1r60
 1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 28.511; Variance 41.538; scale 0.686

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	115	85.2	1475	B33135	gtfB protein precursor	4.19e-11
2	111	82.2	1375	JT0345	dextranucrase (EC 2.3.1.6)	3.42e-10
3	97	71.9	1592	A38175	glucosyltransferase P	4.36e-07
4	92	68.1	1290	JC5473	dextranucrase (EC 2.3.1.6)	5.15e-06
5	86	63.7	1365	A41483	glucosyltransferase (EC 2.3.1.6)	9.30e-05
6	79	58.5	1431	A45866	dextranucrase (EC 2.3.1.6)	2.44e-03
7	78	57.8	575	S46329	intermediate filament	3.85e-03
8	77	57.0	1518	A44811	glucosyltransferase (EC 2.3.1.6)	6.06e-03
9	74	54.8	644	S55395	neurofilament protein	2.32e-02
10	74	54.8	1599	S22737	glucosyltransferase (EC 2.3.1.6)	2.32e-02
11	72	53.3	101	S51349	hypothetical protein	5.58e-02
12	71	52.6	532	QFPGM	neurofilament triplet	8.62e-02
13	71	52.6	845	A45669	neurofilament triplet	8.62e-02
14	71	52.6	916	A27864	neurofilament triplet	8.62e-02
15	70	51.9	494	I52658	neurofilament-66 - hu	1.33e-01
16	70	51.9	504	I53868	alpha-interneuron - hu	1.33e-01
17	68	50.4	505	A41023	alpha-interneuron - ra	1.33e-01
18	68	50.4	372	S28296	hypothetical protein	3.11e-01
19	67	49.6	784	PN0009	neurofilament triplet	3.11e-01
20	67	49.6	798	I50479	neurofilament medium	4.73e-01
21	67	49.6	849	S00030	neurofilament triplet	4.73e-01
22	67	49.6	858	S15762	neurofilament triplet	4.73e-01
23	65	48.1	472	S41720	intermediate filament	1.08e+00

24	65	48.1	1110	2	I51116	NF-180 - sea lamprey	1.08e+00
25	64	47.4	342	2	C48435	cysteine proteinase A	1.63e+00
26	63	46.7	771	2	E70701	probable bioef protei	2.44e+00
27	63	46.7	873	2	S26690	glycoprotein B - infe	2.44e+00
28	63	46.7	883	1	VGBEIL	glycoprotein B precu	2.44e+00
29	63	46.7	883	1	VGBEIL	glycoprotein B precu	2.44e+00
30	62	45.9	132	2	S66478	cytochrome P450 (CYP4	3.64e+00
31	62	45.9	265	2	JC5530	T-cluster binding pro	3.64e+00
32	62	45.9	287	2	B70184	hypothetical protein	3.64e+00
33	62	45.9	375	2	A46174	RNA-binding protein T	3.64e+00
34	62	45.9	386	2	S72435	RNA-binding protein T	3.64e+00
35	62	45.9	392	2	E72436	RNA-binding protein T	3.64e+00
36	61	45.2	290	2	E71256	probable P26 - syphil	5.41e+00
37	61	45.2	430	2	I56572	glial fibrillary acid	5.41e+00
38	61	45.2	432	2	A32936	glial fibrillary acid	5.41e+00
39	61	45.2	448	2	I48128	vimentin - Chinese ha	5.41e+00
40	61	45.2	448	2	S53906	YMK1 protein - yeast	5.41e+00
41	61	45.2	466	2	S22119	vimentin - rat	5.41e+00
42	61	45.2	466	2	A43803	vimentin - mouse	5.41e+00
43	61	45.2	582	2	I48673	matrix metalloprotein	5.41e+00
44	61	45.2	582	2	I38028	matrix metalloprotein	5.41e+00
45	61	45.2	582	2	I84471	matrix metalloprotein	5.41e+00

ALIGNMENTS

RESULT 1
 ENTRY B33135 #type complete
 TITLE gtfB protein precursor - Streptococcus mutans
 ORGANISM #formal_name Streptococcus mutans
 DATE 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 09-Sep-1997

ACCESSIONS
 REFERENCE B33135; A33128
 #authors Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
 #journal J. Bacteriol. (1987) 169:4263-4270
 #cross-references NCBI:87308013
 #accession B33135
 #status Preliminary

REFERENCE B33135; A33128
 #molecule_type DNA
 #residues 1-1475 #label SH1
 #cross-references GB:M17361; NID:G153639; PID:G153640
 #accession A33128
 #authors Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
 #submission submitted to the Protein Sequence Database, September 1990
 #accession A33128
 #status Preliminary; not compared with conceptual translation

CLASSIFICATION #superfamily cpl repeat homology
 FEATURE 1096-1115 #domain cpl repeat homology #label CP1
 1224-1243 #domain cpl repeat homology #label CP2
 1289-1308 #domain cpl repeat homology #label CP3
 1354-1373 #domain cpl repeat homology #label CP4
 1419-1438 #domain cpl repeat homology #label CP5
 #length 1475 #molecular_weight 165811 #checksum 7497

SUMMARY
 Query Match 85.2%; Score 115; DB 2; Length 1475;
 Best local similarity 95.0%; Pred. No. 4.19e-11;
 Matches 19; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Db 552 VPYSFIR-AHDSFVODLIA 570
 QY 1 VPYSFIRTAHDSFVODLIA 20

RESULT 2
 ENTRY JT0345 #type complete
 TITLE dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)

ALTERNATE_NAMES sucrose 6-glucosyltransferase
 ORGANISM #formal_name Streptococcus mutans
 DATE 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 22-Nov-1996

ACCESSIONS JT0345; C33135
 REFERENCE JT0345
 #authors Ueda, S.; Shiroza, T.; Kuramitsu, H.K.
 #journal Gene (1988) 69:101-109
 #title Sequence analysis of the glfC gene from Streptococcus mutans GS-5.

#cross-references MUID:89137980
 #accession JT0345
 #molecule_type DNA
 #residues 1-1375 ##label UED
 #experimental_source GS-5

REFERENCE A33135
 #authors Shiroza, T.; Ueda, S.; Kuramitsu, H.K.
 #journal J. Bacteriol. (1987) 169:4263-4270
 #title Sequence analysis of the glfB gene from Streptococcus mutans.

#cross-references MUID:87308013
 #accession C33135
 #status Preliminary
 #molecule_type DNA
 #residues 1-1349 ##label SHI
 #cross-references GB:M17361

GENETICS
 #gene glfC
 #FUNCTION catalyzes the synthesis of both water-soluble and water-insoluble glucans from glucose.

CLASSIFICATION
 #keywords #superfamily cpl repeat homology
 #duplication; glycosyltransferase; hexosyltransferase; transferase

FEATURE
 1-34 #domain signal sequence #status predicted #label SIG
 35-1375 #product glucosyltransferase #status predicted #label MAT

1126-1145 #domain cpl repeat homology #label CP1\
 1253-1272 #domain cpl repeat homology #label CP2\
 1318-1337 #domain cpl repeat homology #label CP3

SUMMARY #length 1375 #molecular-weight 153021 #checksum 7015

Query Match 82.2%; Score 111; DB 2; Length 1375;
 Best Local Similarity 94.7%; Pred. No. 3,42e-10;
 Matches 18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Db 578 VPYSFIRAHSEVODLI 595
 1 VPYSFIRAHSEVODLI 19

RESULT 3
 ENTRY A38175 #type complete
 TITLE glucosyltransferase precursor - Streptococcus sobrinus
 ORGANISM #formal_name Streptococcus sobrinus
 DATE 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change 09-Sep-1997

ACCESSIONS A38175
 REFERENCE A38175
 #authors Abo, H.; Msumura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.
 #journal J. Bacteriol. (1991) 173:989-996
 #title Peptide sequences for sucrose splitting and glucan binding within Streptococcus sobrinus glucosyltransferase (water-insoluble glucan synthetase).

#cross-references MUID:91123227
 #accession A38175
 #status Preliminary
 #molecule_type DNA
 #residues 1-1592 ##label ABO
 #cross-references GB:D90213; MUID:9217032; PID:dl014946; PID:9217033

CLASSIFICATION #superfamily cpl repeat homology

1093-1112 #domain cpl repeat homology #label CP1\
 1222-1241 #domain cpl repeat homology #label CP2\
 1287-1306 #domain cpl repeat homology #label CP3\
 1330-1351 #domain cpl repeat homology #label CP4\
 1352-1371 #domain cpl repeat homology #label CP5\
 1402-1420 #domain cpl repeat homology #label CP6\
 1465-1484 #domain cpl repeat homology #label CP7\
 1513-1532 #domain cpl repeat homology #label CP8

SUMMARY #length 1592 #molecular-weight 176167 #checksum 5940

Query Match 71.9%; Score 97; DB 2; Length 1592;
 Best Local Similarity 84.2%; Pred. No. 4.36e-07;
 Matches 16; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

Db 548 VPYSFIRAHSEVODLI 565
 1 VPYSFIRAHSEVODLI 19

RESULT 4
 ENTRY JC5473 #type complete
 TITLE dextranucrase (EC 2.4.1.5) - leuconostoc mesenteroides
 ORGANISM #formal_name Leuconostoc mesenteroides
 DATE 07-Jul-1997 #sequence_revision 29-Aug-1997 #text_change 29-Aug-1997

ACCESSIONS JC5473
 REFERENCE JC5473
 #authors Monchois, V.; Willemot, R.M.; Renaud-Simeon, M.; Croux, C.; Monsan, P.
 #journal Gene (1996) 182:23-32
 #title Cloning and sequencing of a gene coding for a novel dextranucrase from Leuconostoc mesenteroides NRRL B-1299 synthesizing only alpha (1-6) and alpha (1-3) linkages.

#cross-references MUID:97136686
 #accession JC5473
 #status nucleic acid sequence not shown; translation not shown
 #molecule_type DNA
 #residues 1-1290 ##label MON
 #cross-references GB:U38181

COMMENT This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose onto acceptor molecules.

GENETICS
 #gene dsra
 #keywords glycosyltransferase; hexosyltransferase
 #FEATURE 78-870
 922-1290 #domain catalytic #status predicted #label CAT\
 #domain glucan-binding #status predicted #label GCB
 #length 1290 #molecular-weight 145572 #checksum 9386

SUMMARY #length 1290 #molecular-weight 145572 #checksum 9386

Query Match 68.1%; Score 92; DB 2; Length 1290;
 Best Local Similarity 78.9%; Pred. No. 5.15e-06;
 Matches 15; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Db 388 PNTSFIRAHSEVODIIA 405
 2 PNTSFIRAHSEVODIIA 20

RESULT 5
 ENTRY A41483 #type complete
 TITLE glucosyltransferase (EC 2.4.1.-) glfs precursor - Streptococcus sobrinus
 ORGANISM #formal_name Streptococcus sobrinus
 DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Sep-1997

ACCESSIONS A41483
 REFERENCE A41483
 #authors Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.
 #journal Infect. Immun. (1990) 58:2452-2458
 #title Analysis of the Streptococcus downei glfs gene, which specifies a glucosyltransferase that synthesizes soluble glucans.

#cross-references MUID:90316665

 WIREIMAGE (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Msrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Tue Jan 11 15:42:29 2000; Maspar time 2.40 Seconds
 Target output not generated. 235,764 Million cell updates/sec

Title: >US-09-290-049-2
 Description: (1-20) from US09290049.pap
 Perfect Score: 135
 Sequence: 1 VPYSFIRAHNSEVODLIA 20

Scoring table: PAM 150
 Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: swiss-prot37
 1:swissprot

Statistics: Mean 29.305; Variance 37.765; scale 0.776

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	115	85.2	1475	1	GLUCOSYLTRANSFERASE-I	9.14e-13
2	111	82.2	1375	1	GLUCOSYLTRANSFERASE-SI	9.56e-10
3	103	76.3	1597	1	GLUCOSYLTRANSFERASE-I	2.79e-08
4	97	71.9	1592	1	GLUCOSYLTRANSFERASE-I	1.09e-05
5	86	63.7	1365	1	GLUCOSYLTRANSFERASE-S	4.10e-04
6	79	58.5	1430	1	GLUCOSYLTRANSFERASE-S	4.97e-03
7	74	54.8	644	1	NEUROFILAMENT TRIPLET	2.12e-02
8	71	52.6	454	1	NEUROFILAMENT TRIPLET	2.12e-02
9	71	52.6	845	1	NEUROFILAMENT TRIPLET	2.12e-02
10	71	52.6	915	1	NEUROFILAMENT TRIPLET	2.12e-02
11	70	51.9	504	1	ALPHA-INTEREXIN (ALPH	3.42e-02
12	70	51.9	505	1	ALPHA-INTEREXIN (ALPH	3.42e-02
13	68	50.4	372	1	HYPOHETICAL 42.4 KD P	8.75e-02
14	67	49.6	848	1	NEUROFILAMENT TRIPLET	1.39e-01
15	67	49.6	857	1	NEUROFILAMENT TRIPLET	1.39e-01
16	64	47.4	465	1	VIMENTIN	5.43e-01
17	63	46.7	873	1	GLYCOPROTEIN B PRECURS	8.48e-01
18	63	46.7	883	1	GLYCOPROTEIN B PRECURS	8.48e-01
19	63	46.7	883	1	GLYCOPROTEIN B PRECURS	8.48e-01
20	62	45.9	375	1	NUCLEOLYSIN TIR-1 (TIA	1.32e+00
21	62	45.9	386	1	NUCLEOLYSIN TIR-1 (TIA	1.32e+00
22	62	45.9	392	1	NUCLEOLYSIN TIR-1 (TIA	1.32e+00
23	61	45.2	428	1	GLIAL FIBRILLARY ACIDI	2.03e+00

RESULT ID	1	STANDARD	PRT	1475 AA	ALIGNMENTS
AC	01-NOV-1988 (REL. 09, CREATED)				
DT	01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)				
DT	15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)				
DE	GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-1) (DEXTRANSUCRASE)				
DE	(SUCROSE 6-GLUCOSYLTRANSFERASE).				
GN	GTFP.				
OS	STREPTOCOCCUS MUTANS.				
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;				
OC	STREPTOCOCCUS.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-GS-5;				
RX	MEDLINE: 87308013.				
RA	SHIROZA T., UEDA S., KURAMITSU H.K.;				
RT	"Sequence analysis of the gtfb gene from Streptococcus mutans.";				
RL	J. BACTERIOL. 169:4263-4270(1987).				
CC	- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.				
CC	- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).				
CC	- D-PRUCOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).				
CC	- SUBCELLULAR LOCATION: SECRETED.				
CC	- DISEASE: DENTAL CARIES.				
CC	- GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-S SYNTHESIZES BOTH FORMS OF GLUCANS.				
CC	- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-BINDING PROTEIN FROM S. MUTANS.				
CC	- THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
DR	EMBL: M17361; G153640; -				
DR	PIR: B33135; B33135				
DR	PFAM: PF00128; alpha-amylase; 1.				
DR	TRANSFERASE; GLUCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.				
FT	SIGNAL	1	34	POTENTIAL.	

FT CHAIN 35 1475 GLUCOSYLTRANSFERASE-I.
FT DOMAIN 35 1050 CATALYTIC (APPROXIMATE).
FT TRANSFERASE 1096 1475 GLUCAN-BINDING (APPROXIMATE).
FT REPEAT 1096 1129 A REPEAT.
FT DOMAIN 1160 1469 5 X TANDEM REPEATS.
FT REPEAT 1160 1209 1.
FT REPEAT 1224 1274 2.
FT REPEAT 1289 1339 3.
FT REPEAT 1354 1404 4.
FT REPEAT 1419 1469 5.
SQ SEQUENCE 1475 AA; 165812 MW; 4542C1D6 CRC32;
Query Match 85.2%; Score 115; DB 1; Length 1475;
Best Local Similarity 95.0%; Pred. No. 9,14e-13;
Matches 19; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
Db 552 VPSYSEIR-AHDEVDLLA 570
1 VPSYSEIRAHDEVDLLA 20
RESULT 2
ID GTFC_STRMU STANDARD; PRT; 1375 AA.
AC P13470; P05427;
DT 01-NOV-1988 (REL. 09, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GLUCOSYLTRANSFERASE-SI PRECURSOR (EC 2.4.1.5) (GTF-SI)
DE (DEXTRANSUCRASE) (SUCROSE 6-GLUCOSYLTRANSFERASE).
GN GTFC.
OS STREPTOCOCCUS MUTANS.
OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-GS-5;
RX MEDLINE: 89137980.
RA UEDA S., SHIROGA T., KURAMITSU H.K.;
RT "Sequence analysis of the gltf gene from Streptococcus mutans GS-5.";
RL GENE 69:101-109(1988).
RN (2)
RP SEQUENCE OF 1-349 FROM N.A.
RC STRAIN-GS-5;
RX MEDLINE: 87308013.
RA SHIROGA T., UEDA S., KURAMITSU H.K.;
RT "Sequence analysis of the gltf gene from Streptococcus mutans.";
RL J. BACTERIOL. 169:4263-4270(1987).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) -
D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DISEASE: DENTAL CARIES.
CC -1- GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED
GLUCOSE AND SOME 1,6 LINKAGES). GTF-S SYNTHESIZES WATER-SOLUBLE
GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF
GLUCANS.
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
BINDING PROTEIN FROM S. MUTANS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL: M2054; G153643;
CC EMBL: M17361; G153641;
CC PIR: J10345; J10345.

DR PIR: C33135; C33135.
DR PFAM: PF00128; alpha-amylase; 1.
KW TRANSFERASE; GLYCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.
FT SIGNAL 1 34
FT CHAIN 35 1375 GLUCOSYLTRANSFERASE-SI.
FT DOMAIN 35 1050 CATALYTIC (APPROXIMATE).
FT TRANSFERASE 1126 1375 GLUCAN-BINDING (APPROXIMATE).
FT REPEAT 1126 1375 2.4 A, 1 C AND 1 AC REPEATS.
FT REPEAT 1169 1159 A REPEAT.
FT REPEAT 1200 1200 A REPEAT.
FT REPEAT 1227 1238 C REPEAT.
FT REPEAT 1253 1303 AC REPEAT.
FT REPEAT 1318 1330 A REPEAT (INCOMPLETE).
SQ SEQUENCE 1375 AA; 153022 MW; 3EA3727E CRC32;
Query Match 82.2%; Score 111; DB 1; Length 1375;
Best Local Similarity 94.7%; Pred. No. 9,54e-12;
Matches 18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
Db 578 VPSYSEIR-AHDEVDLLA 595
1 VPSYSEIRAHDEVDLLA 19
OY 1 VPSYSEIRAHDEVDLLA 19
RESULT 3
ID GTF1_STRDO STANDARD; PRT; 1597 AA.
AC P11001;
DT 01-JUL-1989 (REL. 11, CREATED)
DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GLUCOSYLTRANSFERASE-I PRECURSOR (EC 2.4.1.5) (GTF-1) (DEXTRANSUCRASE)
DE (SUCROSE 6-GLUCOSYLTRANSFERASE).
GN GTF1.
OS STREPTOCOCCUS DOWNEI (STREPTOCOCCUS SOBRINUS).
OC BACTERIA: FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC STREPTOCOCCUS.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-MFE28;
RX MEDLINE: 87308014.
RA FERRETTI J.J., GILPIN M.L., RUSSELL R.B.;
RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus
sibirinus MFE28.";
RL J. BACTERIOL. 169:4271-4278(1987).
CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIANE THE
AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL) (N) -
D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL) (N+1).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DISEASE: DENTAL CARIES.
CC -1- GTF-I SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED
GLUCOSE AND SOME 1,6 LINKAGES). GTF-S SYNTHESIZES WATER-SOLUBLE
GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF
GLUCANS.
CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
BINDING PROTEIN FROM S. MUTANS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL: M17391; G153647;
CC PFAM: PF00128; alpha-amylase; 1.
KW TRANSFERASE; GLYCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.
FT SIGNAL 1 38
FT CHAIN 39 1597 POTENTIAL.
FT DOMAIN 39 1050 GLUCOSYLTRANSFERASE-I.
FT CATALYTIC (APPROXIMATE).

FT DOMAIN 1099 1597 GLUCAN-BINDING (APPROXIMATE).
 FT DOMAIN 1099 1597 1.25 A, 2 B, AND 5 AC REPEATS.
 FT REPEAT 1099 1132 A REPEAT.
 FT REPEAT 1163 1213 AC REPEAT.
 FT REPEAT 1227 1277 AC REPEAT.
 FT REPEAT 1292 1342 AC REPEAT.
 FT REPEAT 1352 1399 B REPEAT.
 FT REPEAT 1406 1455 AC REPEAT.
 FT REPEAT 1465 1512 B REPEAT.
 FT REPEAT 1519 1568 AC REPEAT.
 FT REPEAT 1582 1597 A REPEAT (INCOMPLETE).
 SQ SEQUENCE 1597 AA: 177080 MW: 995233CA CRC32:

Query Match 76.3%; Score 103; DB 1; Length 1597;
 Best Local Similarity 89.5%; Pred. No. 9,56e-10;
 Matches 17; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Db 554 VPSYSFAR-AHSEVODLI 571
 QY 1 VPSYSFIRTAHSEVODLI 19
 ID GTF2_STRDO STANDARD; PRT: 1592 AA.
 AC P27470:
 DT 01-AUG-1992 (REL. 23, CREATED)
 DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE GLUCOSYLTRANSFERASE-1 PRECURSOR (EC 2.4.1.5) (GTF-1) (DEXTRANSUCRASE)
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).
 OS STREPTOCOCCUS DOWNEI (STREPTOCOCCUS SOBRINUS).
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
 OC STREPTOCOCCUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-6715:
 RX MEDLINE: 91123227.
 RA ABO H., MATSUMURA T., KODAMA T., OHTA H., FUKUI K., KATO K.,
 RA KAGAWA H.;
 RT "Peptide sequences for sucrose splitting and glucan binding within
 RT streptococcus sobrinus glucosyltransferase (water-insoluble glucan
 RT synthetase).";
 RT J. BACTERIOL. 173:989-996(1991).
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
 CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
 CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
 CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
 CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -
 CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- DISEASE: DENTAL CARIES.
 CC -1- GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED
 CC GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES BOTH FORMS OF
 CC GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF
 CC GLUCANS.
 CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
 CC BINDING PROTEIN FROM S. MUTANS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D90213; G217033; -
 CC DR PIR: A38175; A38175.
 CC DR PFAM: PF00128; alpha-amyase; 1.
 CC DR HSSP: P00695; 2HEE.
 CC KW TRANSFERASE; GLUCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.
 CC FT SIGNAL 1 38 POTENTIAL.
 CC FT CHAIN 39 1592 GLUCOSYLTRANSFERASE-1.

FT DOMAIN 39 1044 CATALYTIC (APPROXIMATE).
 FT DOMAIN 1093 1592 GLUCAN-BINDING (APPROXIMATE).
 FT REPEAT 1093 1592 6-5 X TANDEM REPEATS.
 FT REPEAT 1093 1142 1.
 FT REPEAT 1158 1207 2.
 FT REPEAT 1222 1272 3.
 FT REPEAT 1287 1337 4.
 FT REPEAT 1402 1451 5.
 FT REPEAT 1514 1563 6.
 FT REPEAT 1577 1592 7 (INCOMPLETE).
 SQ SEQUENCE 1592 AA: 176167 MW: 273233FA CRC32:

Query Match 71.9%; Score 97; DB 1; Length 1592;
 Best Local Similarity 84.2%; Pred. No. 2,79e-08;
 Matches 16; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

Db 548 VPSYSFAR-AHSEVODLI 565
 QY 1 VPSYSFIRTAHSEVODLI 19
 ID GTF5_STRDO STANDARD; PRT: 1365 AA.
 AC P29336:
 DT 01-DEC-1992 (REL. 24, CREATED)
 DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
 DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
 DE GLUCOSYLTRANSFERASE-5 PRECURSOR (EC 2.4.1.5) (GTF-S) (DEXTRANSUCRASE)
 DE (SUCROSE 6-GLUCOSYLTRANSFERASE).
 OS STREPTOCOCCUS DOWNEI (STREPTOCOCCUS SOBRINUS).
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
 OC STREPTOCOCCUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ME28;
 RX MEDLINE: 90316665.
 RA GILMORE K.S., RUSSELL R.R., FERRETTI J.J.;
 RT "Analysis of the Streptococcus downei gtf5 gene, which specifies a
 RT glucosyltransferase that synthesizes soluble glucans.";
 RT INFECT. IMMUN. 58:2452-2458(1990).
 CC -1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT
 CC TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE
 CC OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE
 CC AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.
 CC -1- CATALYTIC ACTIVITY: SUCROSE + (1,6-ALPHA-D-GLUCOSYL)(N) -
 CC D-FRUCTOSE + (1,6-ALPHA-D-GLUCOSYL)(N+1).
 CC -1- ENZYME REGULATION: GLUCAN SYNTHESIS BY GTF-S IS INDEPENDENT OF
 CC PRIMER GLUCAN UNLIKE GTF-1.
 CC -1- DISEASE: DENTAL CARIES.
 CC -1- GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE).
 CC -1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-
 CC BINDING PROTEIN FROM S. MUTANS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M30943; G153653; -
 CC DR PIR: A1483; A1483.
 CC DR PFAM: PF00128; alpha-amyase; 1.
 CC KW TRANSFERASE; GLUCOSYLTRANSFERASE; SIGNAL; REPEAT; DENTAL CARIES.
 CC FT SIGNAL 1 36 OR 37 (POTENTIAL).
 CC FT CHAIN 37 1365 GLUCOSYLTRANSFERASE-S.
 CC FT DOMAIN 1083 1365 CATALYTIC (APPROXIMATE).
 CC FT DOMAIN 1083 1365 GLUCAN-BINDING (APPROXIMATE).
 CC FT REPEAT 1083 1131 4.5 X TANDEM REPEATS.
 CC FT REPEAT 1150 1199 2.

DE NEUROFILAMENT TRIPLET M PROTEIN (160 KD NEUROFILAMENT PROTEIN) (NF-M)
 DE (FRAGMENT).
 GN NEFM.
 OS SUS SCROFA (PIG).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.
 RN [1]
 RP SEQUENCE.
 RC TISSUE-SPINAL CORD;
 RX MEDLINE: 85076594.
 RA GEISLER N., FISCHER S., VANDEKERCKHOVE J., PLESSMANN U., WEBER K.;
 RT "Hybrid character of a large neurofilament protein (NF-M):
 RT intermediate filament type sequence followed by a long and acidic
 RT carboxy-terminal extension.";
 RT EMBO J. 3:2701-2706(1984).
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPLET K-S-P, NFM IS
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 CC THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 CC OF AXONAL CALIBER.
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC PIR: A05075; OEPGM.
 DR PROSITE: PS00226; IF: 1.
 DR PFAM: PF00038; filament; 1.
 DR INTERMEDIATE FILAMENT; HEPTAD REPEAT PATTERN; COILED COIL; NEURONE;
 KW ACETYLATION; PHOSPHORYLATION; GLYCOPROTEIN.
 FT MOD_RES 1 1
 FT DOMAIN 1 104 HEAD.
 FT FT 105 412 ROD.
 FT FT 105 >454 TAIL.
 FT FT 105 136 COIL 1A.
 FT FT 137 149 LINKER 1.
 FT FT 150 248 COIL 1B.
 FT FT 249 265 LINKER 12.
 FT FT 266 287 COIL 2A.
 FT FT 288 291 LINKER 2.
 FT FT 292 411 COIL 2B.
 FT FT 411 417 GLCNAC (BY SIMILARITY).
 FT CARBOHYD 432 432 GLCNAC (BY SIMILARITY).
 FT NON_TER 454 454
 SQ SEQUENCE 454 AA; 51854 MW; 3DFC58A1 CRC32;
 Query Match 52.6%; Score 71; DB 1; Length 454;
 Best Local Similarity 50.0%; Pred. No. 2,12e-02;
 Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
 Db 232 AFLRSNHEEVADLLA 247
 QY 5 SFIRTAHSEVODLLA 20

RT the middle molecular weight neurofilament protein.";
 RL J. NEUROSCI. 7:2590-2599(1987).
 RN [2]
 RP PHOSPHORYLATION SITES, AND REVISION TO 500.
 RX MEDLINE: 92165797.
 RA XU Z.-S., LIU W.-S., WILLARD M.B.;
 RT "Identification of six phosphorylation sites in the COOH-terminal
 RT tail region of the rat neurofilament protein M.";
 RL J. BIOL. CHEM. 267:4467-4471(1992).
 RN [3]
 RP CARBOHYDRATE-BINDING SITES.
 RX MEDLINE: 93346421.
 RA DONG D.L.-Y., XU Z.-S., CHEVRIER M.R., COTTER R.J., CLEVELAND D.W.,
 RA HART G.W.;
 RT "Glycosylation of mammalian neurofilaments. Localization of multiple
 RT O-linked N-acetylglucosamine moieties on neurofilament polypeptides
 RT L and M.";
 RL J. BIOL. CHEM. 268:16679-16687(1993).
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPLET K-S-P, NFM IS
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 CC THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 CC OF AXONAL CALIBER.
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: M18628; G205688; -
 DR PROSITE: PS00226; IF: 1.
 DR PFAM: PF00038; filament; 1.
 DR INTERMEDIATE FILAMENT; HEPTAD REPEAT PATTERN; COILED COIL; NEURONE;
 KW PHOSPHORYLATION; GLYCOPROTEIN.
 FT INIT_MET 0 0
 FT DOMAIN 1 103 HEAD.
 FT FT 103 410 ROD.
 FT FT 103 410 TAIL.
 FT DOMAIN 103 134 COIL 1A.
 FT FT 135 147 LINKER 1.
 FT FT 148 246 COIL 1B.
 FT DOMAIN 247 263 LINKER 12.
 FT FT 264 285 COIL 2A.
 FT FT 286 289 LINKER 2.
 FT DOMAIN 290 410 COIL 2B.
 FT FT 410 417 GLCNAC.
 FT CARBOHYD 430 430 GLCNAC.
 FT MOD_RES 502 502 PHOSPHORYLATION.
 FT MOD_RES 506 506 PHOSPHORYLATION.
 FT MOD_RES 536 536 PHOSPHORYLATION.
 FT MOD_RES 603 603 PHOSPHORYLATION.
 FT MOD_RES 608 608 PHOSPHORYLATION.
 FT MOD_RES 666 666 PHOSPHORYLATION.
 FT CONFLICT 500 500 MISSING (IN REF. 1).
 SQ SEQUENCE 845 AA; 95660 MW; 8C2B516C CRC32;
 Query Match 52.6%; Score 71; DB 1; Length 845;
 Best Local Similarity 50.0%; Pred. No. 2,12e-02;
 Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
 Db 230 AFLRSNHEEVADLLA 245
 QY 5 SFIRTAHSEVODLLA 20

RA NAPOLITANO E.W., CHIN S.S.M., COLMAN D.R., LIEM R.K.H.;
 "Complete amino acid sequence and in vitro expression of rat NF-M,"


```

CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X05640; G297529;
DR EMBL: M20481; G300040;
DR PIR: S00030; S00030.
DR MGD: MG1:97314; NFM.
DR PROSITE: PS00226; IF; 1.
DR PIR: P00038; filament; 1.
CC INTERMEDIATE FILAMENT: HEPTAD REPEAT PATTERN; COILED COIL; NEURONE;
CC PHOSPHORYLATION; GLYCOPROTEIN.
CC -----
FT INIT_MET 0 0
FT DOMAIN 1 102 HEAD.
FT DOMAIN 103 410 ROD.
FT DOMAIN 411 848 TAIL.
FT DOMAIN 103 134 COIL 1A.
FT DOMAIN 135 147 LINKER 1.
FT DOMAIN 148 246 COIL 1B.
FT DOMAIN 247 263 LINKER 12.
FT DOMAIN 264 285 COIL 2A.
FT DOMAIN 286 289 LINKER 2.
FT DOMAIN 290 410 COIL 2B.
FT CARBOHYD 47 47 GLCNAC (BY SIMILARITY).
FT CARBOHYD 430 430 GLCNAC (BY SIMILARITY).
FT CONFLICT 432 432 S -> F (IN REF. 2).
FT CONFLICT 539 540 QA -> RR (IN REF. 2).
SQ SEQUENCE 848 AA; 95910 MW; E06A637A CRC32;

Query Match 49.6%; Score 67; DB 1; Length 848;
Best Local Similarity 50.0%; Pred. No. 1.39e-01;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

DB 230 AFLRNHEEVADLLA 245
:|:|:|:|:|:|
5 SFRTADSEVODLIA 20

RESULT 15
ID NFM_CHICK STANDARD; PRT; 857 AA.
AC P16053;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE NEUROFILAMENT TRIPLET M PROTEIN (160 KD NEUROFILAMENT PROTEIN) (NF-M).
GN NFM.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90174973.
RA ZOPF D., DINEVA B., BETZ H., GUNDELINGER E.D.;
RT "Isolation of the chicken middle-molecular weight neurofilament
RT (NF-M) gene and characterization of its promoter.";
RL NUCLEIC ACIDS RES. 18:521-529(1990).
RN [2]
RP SEQUENCE OF 259-857 FROM N.A.
RX MEDLINE: 88112814.
RA ZOPF D., HERMANS-BORGMEYER I., GUNDELINGER E.D., BETZ H.;
RT "Identification of gene products expressed in the developing chick

```

```

RT visual system: characterization of a middle-molecular-weight
RT neurifilament cDNA.";
RL GENES DEV. 1:699-708(1987).
CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPLETTE K-S-P, NFM IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X17102; G63689;
DR EMBL: X05558; G63686;
DR PIR: A27040; A27040.
DR PIR: S08061; S08061.
DR PIR: S15762; S15762.
DR PROSITE: PS00226; IF; 1.
DR PIR: P00038; filament; 1.
CC INTERMEDIATE FILAMENT: HEPTAD REPEAT PATTERN; COILED COIL; NEURONE;
CC PHOSPHORYLATION; GLYCOPROTEIN.
CC -----
FT INIT_MET 0 0
FT DOMAIN 1 98 HEAD.
FT DOMAIN 99 406 ROD.
FT DOMAIN 407 857 TAIL.
FT DOMAIN 99 130 COIL 1A.
FT DOMAIN 131 143 LINKER 1.
FT DOMAIN 144 242 COIL 1B.
FT DOMAIN 243 259 LINKER 12.
FT DOMAIN 260 281 COIL 2A.
FT DOMAIN 282 285 LINKER 2.
FT DOMAIN 286 406 COIL 2B.
FT CARBOHYD 46 46 GLCNAC (BY SIMILARITY).
FT CARBOHYD 426 426 GLCNAC (BY SIMILARITY).
FT CONFLICT 546 546 G -> R (IN REF. 2).
SQ SEQUENCE 857 AA; 95704 MW; 3D05FDD CRC32;

Query Match 49.6%; Score 67; DB 1; Length 857;
Best Local Similarity 43.8%; Pred. No. 1.39e-01;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

DB 226 AFLRNHEEVADLLA 241
:|:|:|:|:|:|
5 SFRTADSEVODLIA 20

```

Search completed: Tue Jan 11 15:42:37 2000
Job time : 8 secs.

 WISE (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Msrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Tue Jan 11 15:36:00 2000; Maspar time 5.50 Seconds

81.178 Million cell updates/sec

Output not generated.

Title: >US-09-290-049-1

Description: (1-21) from US09290049.pep

Perfect Score: 148

Sequence: 1 ANDHLSILEAMSDNDPYLHD 21

Scoring table: PAM 150

Gap 15

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

a-geneseq35
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
 14:part14 15:part15 16:part16 17:part17 18:part18
 19:part19 20:part20 21:part21 22:part22 23:part23
 24:part24 25:part25 26:part26 27:part27 28:part28
 29:part29 30:part30 31:part31 32:part32 33:part33
 34:part34 35:part35 36:part36 37:part37 38:part38
 39:part39

Statistics: Mean 20.949; Variance 72.779; scale 0.288

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	131	88.5	1592	6 R32925	Glucosyltransferase	5.55e-06
2	83	56.1	1577	16 R91047	Alpha-D-glucosyltrans	6.09e-01
3	66	44.6	387	33 W62297	Scenedesmus D1 protea	2.73e+01
4	63	42.6	388	33 W62298	Wheat D1 protease.	5.20e+01
5	62	41.9	648	22 W17045	Mouse c-raf 1 protein	6.43e+01
6	62	41.9	648	22 W17044	Human c-raf 1 protein	6.43e+01
7	62	41.9	648	22 W17046	Mutant mouse c-raf 1	6.43e+01
8	62	41.9	648	22 W17047	Mutant mouse c-raf 1	6.43e+01
9	62	41.9	648	4 R22560	Mouse mutant c-raf-1.	6.43e+01
10	62	41.9	648	4 R22562	Mouse mutant c-raf-1.	6.43e+01
11	62	41.9	648	4 R22561	Mouse mutant c-raf-1.	6.43e+01
12	62	41.9	648	4 R25277	Human c-raf-1.	6.43e+01
13	62	41.9	648	21 W13107	Human Raf-1.	6.43e+01
14	62	41.9	648	22 W17048	Mutant mouse c-raf 1	6.43e+01
15	62	41.9	648	22 W17049	Mutant mouse c-raf 1	6.43e+01
16	62	41.9	648	32 W62220	Raf-1 protein.	6.43e+01

17	62	41.9	648	4 R22563	Mouse mutant c-raf-1.	6.43e+01
18	62	41.9	648	18 R22559	Mouse c-raf-1.	6.43e+01
19	62	41.9	648	18 R98215	Human Raf1 Kinase.	6.43e+01
20	60	40.5	210	3 R15613	Yeast alpha 2 mating	9.81e+01
21	60	40.5	236	38 W90021	Expressed antigen for	9.81e+01
22	60	40.5	239	38 W89935	Antigen 1 from Cluste	9.81e+01
23	60	40.5	298	34 W71486	Helicobacter polypept	9.81e+01
24	60	40.5	473	33 W62300	Tobacco D1 protease.	9.81e+01
25	59	39.9	178	1 P90061	Human Fc gamma recept	1.21e+02
26	59	39.9	338	12 R65965	T. niyeum GAPDH.	1.21e+02
27	59	39.9	576	17 R85891	WD-40 domain-contg. p	1.21e+02
28	58	39.9	755	59 W81365	Human prothomone conv	1.21e+02
29	58	39.9	67	31 W28316	Staphylococcus aureus	1.49e+02
30	58	39.2	158	11 R56770	Fibrin binding peptid	1.49e+02
31	58	39.2	159	28 W45492	Targeting ligand derl	1.49e+02
32	58	39.2	167	14 R78615	MPD-1 extracellular r	1.49e+02
33	58	39.2	187	14 R47801	Marine programmed cel	1.49e+02
34	58	39.2	288	14 R78614	MPD-1.	1.49e+02
35	58	39.2	392	39 W67891	Human secreted protei	1.49e+02
36	58	39.2	596	14 R78616	Expression vector pME	1.49e+02
37	58	39.2	724	20 W10730	Human Factor XIIIa re	1.49e+02
38	58	39.2	731	3 R14376	Factor XIII subunit a	1.49e+02
39	58	39.2	732	5 R25385	Human Factor XIII.	1.49e+02
40	58	39.2	732	2 P70293	Sequence of human fac	1.49e+02
41	58	39.2	761	1 P82920	A and A' subunits of	1.49e+02
42	58	39.2	771	15 R80495	Japanese oyster trans	1.49e+02
43	58	39.2	783	15 W64560	Candida albicans o1g	1.49e+02
44	57	38.5	222	29 W55287	H. pylori ORF 13ep120	1.83e+02
45	57	38.5	404	30 W55726	H. pylori ORF hp8e100	1.83e+02

ALIGNMENTS

RESULT	ID	Score	Query Match	Length	ID	Description	Pred. No.
1	R32925	standard; Protein: 1592 AA.	88.5%	1592	6	Glucosyltransferase	5.55e-06
2	R91047	standard; Protein: 1577 AA.	56.1%	1577	16	Alpha-D-glucosyltrans	6.09e-01
3	W62297	Scenedesmus D1 protea	44.6%	387	33	Scenedesmus D1 protea	2.73e+01
4	W62298	Wheat D1 protease.	42.6%	388	33	Wheat D1 protease.	5.20e+01
5	W17045	Mouse c-raf 1 protein	41.9%	648	22	Mouse c-raf 1 protein	6.43e+01
6	W17044	Human c-raf 1 protein	41.9%	648	22	Human c-raf 1 protein	6.43e+01
7	W17046	Mutant mouse c-raf 1	41.9%	648	22	Mutant mouse c-raf 1	6.43e+01
8	W17047	Mutant mouse c-raf 1	41.9%	648	22	Mutant mouse c-raf 1	6.43e+01
9	R22560	Mouse mutant c-raf-1.	41.9%	648	4	Mouse mutant c-raf-1.	6.43e+01
10	R22562	Mouse mutant c-raf-1.	41.9%	648	4	Mouse mutant c-raf-1.	6.43e+01
11	R22561	Mouse mutant c-raf-1.	41.9%	648	4	Mouse mutant c-raf-1.	6.43e+01
12	R25277	Human c-raf-1.	41.9%	648	4	Human c-raf-1.	6.43e+01
13	W13107	Human Raf-1.	41.9%	648	21	Human Raf-1.	6.43e+01
14	W17048	Mutant mouse c-raf 1	41.9%	648	22	Mutant mouse c-raf 1	6.43e+01
15	W17049	Mutant mouse c-raf 1	41.9%	648	22	Mutant mouse c-raf 1	6.43e+01
16	W62220	Raf-1 protein.	41.9%	648	32	Raf-1 protein.	6.43e+01

AC R91047;
 DT 22-MAY-1996 (first entry)
 DE Alpha-D-glucosyltransferase.
 KW Alpha-D-glucosyltransferase; primer-independent; soluble glucan;
 KW sucrose; transglucosyltransferase; cloning; Escherichia coli;
 KW phage lambda-C13; vector; plasmid pGS6501; plasmid pGS6502;
 KW gene transfer; crop improvement; storage carbohydrate; pasture;
 KW feedstuff; senescence; dextran; binder; food; pharmaceutical.
 OS Streptococcus salivarius strain ATCC 25975.
 PN M09606173-A1.
 PD 29-FEB-1996.
 PF 24-AUG-1995: AU0527.
 PR 24-AUG-1994: AU-007643.
 PA (GIRF/) GIFFARD P M.
 PA (JACO/) JACQUES N A.
 PA (SIMP/) SIMPSON C L.
 PI Giffard PM, Jacques NA, Simpson CL;
 DP WPI: 96-151376/15.
 N-PSDB: T13139.
 CC Plants contg. new bacterial DNA encoding glucosyl transferase
 PT activity - retain higher levels of stored carbohydrate(s) in a form
 PS readily digestible by ruminants
 PS Claim 4: Page 16-20: 31pp: English.
 CC The sequence represents an alpha-D-glucosyltransferase from
 CC Streptococcus salivarius. The enzyme is primer-independent, and
 CC produces soluble glucan from sucrose. A gene encoding the enzyme
 CC may be cloned and expressed in Escherichia coli using a subclone
 CC of phage lambda-C13, e.g. plasmid pGS6501 or plasmid pGS6502. The
 CC DNA may also be expressed in a transgenic plant, to improve the
 CC level of stored carbohydrate in a pasture plant which normally
 CC contains low levels, or to prevent degradation of stored carbohydrate
 CC during plant senescence. Dextran may be isolated from the plant, for
 CC use as a food binder or pharmaceutical additive. Primer independence
 CC ensures that the enzyme will be functional in plants. The glucan is
 CC poorly degraded in plants but easily degraded by bacteria in the rumen
 CC of grazing livestock.
 CC Sequence 1577 AA;
 SQ

Query Match 56.1%; Score 83; DB 16; Length 1577;
 Best Local Similarity 72.2%; Pred. No. 6.09e+01;
 Matches 13; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

DB 588 alahlsllawsvndhyg 605
 QY 1 ANDHLSTLEAWSDNDTPY 18

AC W62297;
 DT 28-SEP-1998 (first entry)
 DE Scenedesmus D1 protease.
 KW Scenedesmus; D1 protease; herbicide; inhibition; alga; wheat;
 KM detection; identification.
 OS Scenedesmus obliquus.
 PN M09824934-A2.
 PD 11-JUN-1998.
 PF 03-DEC-1997: U21964.
 PR 05-DEC-1996: US-759581.
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 PA Chisholm DA, Diner BA, Donaldson GK, Hershey HP,
 PI Jordan DB, Tang XS, Trost JT, Wang S, Warren PV;
 DP WPI: 98-333349/29.
 N-PSDB: V39818.
 CC New isolated plant D1 protease(s) - used for identifying herbicidal
 PT agents which target the D1 protease by inhibiting the enzyme
 PS Claim 23: Page 44-45: 73pp: English.
 CC A method has been developed for identifying a herbicidal agent which
 CC inhibits D1 protease. The method comprises: (a) incubating a D1 protease
 CC enzyme in a sample suspected of containing a herbicidal agent with a
 CC suitable D1 enzyme substrate, where an enzyme product is formed; and (b)
 CC detecting and quantifying the enzyme product of step (a). Also described

CC In the present invention are: (1) a method for detecting in vivo a
 CC herbicidal agent which inhibits D1 protease comprising: (a) preparing a
 CC reaction mixture comprising: (1) a wild type cell comprising an active D1
 CC protease enzyme capable of processing a D1 pre-protein; and a photosystem
 CC II core complex capable of variable chlorophyll fluorescence; (11) a
 CC suspected herbicidal agent which inhibits D1 protease; and (111) a suitable
 CC growth medium; (b) illuminating the reaction mixture at illumination
 CC conditions to permit D1 turnover; and (c) measuring the variable
 CC chlorophyll fluorescence produced in step (b), where the level of the
 CC variable chlorophyll fluorescence correlates with herbicidal activity of
 CC the suspected herbicidal agent; (2) a gene encoding a D1 protease enzyme
 CC isolated from: (a) wheat; (b) tobacco; (c) spinach; (d) Scenedesmus; or
 CC (e) Synechocystis; (see V39818 to V39822, which encode W62297 to W62301
 CC respectively). Native or recombinant D1 proteases can be used to develop
 CC assays for the detection of herbicidal compositions capable of
 CC inhibiting D1.
 CC Sequence 387 AA;
 SQ

Query Match 44.6%; Score 66; DB 33; Length 387;
 Best Local Similarity 36.8%; Pred. No. 2.73e+01;
 Matches 7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

DB 2 tseqllflawrvdrayv 20
 QY 1 ANDHLSTLEAWSDNDTPYL 19

AC W62298;
 DT 28-SEP-1998 (first entry)
 DE Wheat D1 protease.
 KW Scenedesmus; D1 protease; herbicide; inhibition; alga; wheat;
 KM detection; identification.
 OS Triticum sp.
 PN M09824934-A2.
 PD 11-JUN-1998.
 PF 03-DEC-1997: U21964.
 PR 05-DEC-1996: US-759581.
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 PA Chisholm DA, Diner BA, Donaldson GK, Hershey HP,
 PI Jordan DB, Tang XS, Trost JT, Wang S, Warren PV;
 DP WPI: 98-333349/29.
 N-PSDB: V39819.
 CC New isolated plant D1 protease(s) - used for identifying herbicidal
 PT agents which target the D1 protease by inhibiting the enzyme
 PS Claim 23: Page 47-48: 73pp: English.
 CC A method has been developed for identifying a herbicidal agent which
 CC inhibits D1 protease. The method comprises: (a) incubating a D1 protease
 CC enzyme in a sample suspected of containing a herbicidal agent with a
 CC suitable D1 enzyme substrate, where an enzyme product is formed; and (b)
 CC detecting and quantifying the enzyme product of step (a). Also described
 CC in the present invention are: (1) a method for detecting in vivo a
 CC herbicidal agent which inhibits D1 protease comprising: (a) preparing a
 CC reaction mixture comprising: (1) a wild type cell comprising an active D1
 CC protease enzyme capable of processing a D1 pre-protein; and a photosystem
 CC II core complex capable of variable chlorophyll fluorescence; (11) a
 CC suspected herbicidal agent which inhibits D1 protease; and (111) a suitable
 CC growth medium; (b) illuminating the reaction mixture at illumination
 CC conditions to permit D1 turnover; and (c) measuring the variable
 CC chlorophyll fluorescence produced in step (b), where the level of the
 CC variable chlorophyll fluorescence correlates with herbicidal activity of
 CC the suspected herbicidal agent; (2) a gene encoding a D1 protease enzyme
 CC isolated from: (a) wheat; (b) tobacco; (c) spinach; (d) Scenedesmus; or
 CC (e) Synechocystis; (see V39818 to V39822, which encode W62297 to W62301
 CC respectively). Native or recombinant D1 proteases can be used to develop
 CC assays for the detection of herbicidal compositions capable of
 CC inhibiting D1.
 CC Sequence 388 AA;
 SQ

Query Match 42.6%; Score 63; DB 33; Length 388;
 Best Local Similarity 38.9%; Pred. No. 5.20e+01;

PT Classification of lymphoma or lung cancer - on the basis of a point mutation in c-rat-1 gene
PS Example 3, Column 17-20; 26pp; English.
CC W17046-W17049 are mutant versions of the mouse c-rat-1 protein, derived from four different types of tumour. They were used in a method for classifying a lymphoma or lung cancer, by comparison with the wild-type c-rat-1 protein sequence derived from different tumours in mice. The method, when used for diagnosing human tumours, involves detecting the presence of a point mutation in a conserved region of the c-rat-1 gene (codon 533, encoding Ser) derived from lymphoma or lung cancer tissue CC and classifying the lymphoma or lung cancer as a c-rat-1 mutation-associated cancer if one or more point mutations are present. The CC method is particularly applicable to diagnosis of lung adenocarcinoma. SQ Sequence 648 AA;

```

Query Match      41.9%   Score 62;  DB 22; length 648;
Best Local Similarity 30.0%   Pred. No. 6.43e+01;
Matches      6;  Conservative      7;  Mismatches      7;  Indels      0;  Gaps      0;
413 tkdnlavtgcgcslykh 432
:::|:|:|:|:|:|
1 ANDRLSLEAMSDDNDPIYLH 20

```

RESULT	9	standard; Protein; 648 AA.
ID	R22560	
AC	R22560:	
DT	25-AUG-1992	(first entry)
DE	Mouse mutant c-rat-1.	
OS	Polymerase chain reaction; tumour; cancer; detection; mutation.	
KM	Mus musculus.	
FN	Key	
FT	region	location/Qualifiers
FT	region	62..194
FT	region	/note="conserved region (CR) 1"
FT	region	235..269
FT	region	/note="conserved region (CR) 2"
FT	region	338..627
FT	region	/note="conserved region (CR) 3"
FT	region	513..515
FT	region	/note="APE site"
FT	region	517
FT	region	/note="mutation, Ile -> Val"
PN	US759738-A.	
PD	25-FEB-1992.	
PE	16-SEP-1991.	759738.
PR	16-SEP-1988.	US-236947.
PR	16-SEP-1991.	US-759738.
	(US98)	US DEPT HEALTH & HUMAN.

DR WPI: 92-123681/15.
PT Detecting C-Raf-1 genes - by amplifying region of C-Raf-1 gene,
PT analysing prods. and classifying individual having mutation(s) in
PT region
PS Disclosure: Page 28: 65pp. English.
CC The sequence is that of a mutated version of mouse c-ras-1, the
CC mutation occurs just downstream of the APE site, the mutation is
CC not affectual. The region in which it occurs overlaps an epitope
CC shared by monoclonal antibodies generated against raf (Kotich et al
CC 1990). This region is a hydrophilic domain, the structure of which
CC is predicted to be altered by this mutation.
CC See also R25277 and R22559-R22563.
CC Sequence 648 AA;
CQ

Query Match	41.98;	Score 62;	DB 4;	Length 648;
Best Local Similarity	30.08;	Pred. No. 6.43e+01;		
Matches	6;	Conservative	7;	Mismatches 7;
				Indels 0;
				Gaps 0;

```
Db      413 tkdnlaivtqcwcegsslykh 432
      ::||:|:| :|:|
QY      1 ANDHLSILEAWSNDNDTPYLH 20
```

RESULT 1.6

ID	R22562;	strand; Protein; 648 AA.
AC	R22562;	
DT	25-AUG-1992	(first entry)
DE	Mouse mutant C-raf-1.	
KM	Polymerase chain reaction; tumour; cancer; detection; mutation.	
OS	Mus musculus.	
FH	Key	Location/Qualifiers
FT	region	62..194
FT	region	/note="conserved region (CR) 1"
FT	region	253..269
FT	region	/note="conserved region (CR) 2"
FT	region	338..627
FT	region	/note="conserved region (CR) 3"
FT	region	513..515
FT	region	/note="APE site"
FT	region	527
FT	region	/note="mutation, Phe -> Ser"
PN	US759738-A.	
PD	25-FEB-1992.	
PF	16-SEP-1991; 759738.	
PR	26-AUG-1988; US-236947.	
PR	16-SEP-1991; US-759738.	
PA	(USSH) US DEPT HEALTH & HUMAN.	
PI	Rapp U, Storm S;	
PI	WPI; 92-123681/15.	
PT	Detecting C-Raf-1 genes - by amplifying region of C-Raf-1 gene,	
PT	analysis prods. and classifying individual having mutation(s) in	
PT	region	
PS	Disclosure: Page 35; 65pp; English.	
CC	The sequence is that of a mutated version of mouse C-raf-1, the	
CC	mutation occurs just downstream of the APE site, the mutation is	
CC	not artifactual. The region in which it occurs overlaps an epitope	
CC	shared by monoclonal antibodies generated against raf (Kolch et al	
CC	1990). This region is a hydrophilic domain, the structure of which	
CC	is predicted to be altered by this mutation.	
CC	See also R25277 and R22559-R22563.	
QJ	Sequence 648 AA;	

	Query Match	41.9%	Score 62;	DB 4;	Length 648;
	Best Local Similarity	30.0%;	Pred. No. 6.43e+01;		
Matches	6;	Conservative	7;	Mismatches 7;	Indels 0;
Gaps					0;
Db	413 tkdhalvtgwcgsslykh	432			
	::: :: :: :: :: ::				
Oy	.1 ANDHSLILEAWSDNDREPLH	20			

```

RESULT      11
ID          R23561 standard; Protein; 648 AA.
AC          R23561;
DC          25-AUG-1992 (first entry)
DE          Mouse mutant C-rai-1.
KW          Polymerase chain reaction; tumour; cancer; detection; mutation
OS          Mus musculus.
FH          Key
FT          region
FT          location/Qualifiers
FT          62..194
FT          /note="conserved region (CR) 1"
FT          253..269
FT          /note="conserved region (CR) 2"
FT          338..627
FT          /note="conserved region (CR) 3"
FT          513..515
FT          /note="APE site"
FT          522
FT          region
FT          /note="mutation, Asp -> Asn"
FN          US7759738-A.
PD          25-FEB-1992.
PF          16-SEP-1991; 759738.
PR          26-AUG-1988; US-236947.
PR          16-SEP-1991; US-759738.
PA          (USSH ) US DEPT HEALTH & HUMAN.
PI          RAPP U, Storm S;
DR          WPI; 92-123661/15.

```



```

PT Detecting C-Raf-1 genes - by amplifying region of C-Raf-1 gene,
PT analysing prods. and classifying individual having mutation(s) in
PT region
PS Disclosure: Page 31: 65pp: English.
CC The sequence is that of a mutated version of mouse c-raf-1, the
CC mutation occurs just downstream of the APE site, the mutation is
CC not attetual. The region in which it occurs overlaps an epitope
CC shared by monoclonal antibodies generated against raf (Kolich et al.
CC 1990). This region is a hydrophilic domain, the structure of which
CC is predicted to be altered by this mutation.
CC See also R25277 and R22559-R22565.
SC Sequence 648 AA;
SQ
Query Match 41.9%; Score 62; DB 4; Length 648;
Best Local Similarity 30.0%; Pred. No. 6.43e+01;
Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;
Db 413 tkdnlatvqcwcegslykh 432
:::|:|:|:|:|:|
1 ANDHLSLEAWSNDNDTPYLH 20
R25277 standard; Protein; 648 AA.
R25277;
DT 25-AUG-1992 (first entry)
DE Human c-raf-1.
KW Polymerase chain reaction; tumour; cancer; detection.
OS Homo sapiens.
PN US7759738-A.
PD 25-SEP-1992.
PE 16-SEP-1991; 759738.
PR 26-AUG-1988; US-236947.
PR 16-SEP-1991; US-759738.
PR (USSH ) US DEPT HEALTH & HUMAN.
PI Rapp U, Storm S;
PI WPI: 92-123681/15.
DR N-PSDB: Q23459.
DR Detecting C-Raf-1 genes - by amplifying region of C-Raf-1 gene,
PT analysing prods. and classifying individual having mutation(s) in
PT region
PS Disclosure: Page 46: 65pp: English.
CC The sequence is that of human c-raf-1, the c-raf-1 gene can be used in
CC methods for detecting and treating a wide range of cancers including
CC lung cancer, T-cell lymphomas, renal cell carcinoma, ovarian carcinoma
CC and mixed parotid gland tumours. The methods consist of amplifying
CC the gene by PCR and then, based on the presence or absence of one or
CC more mutations, individuals at increased risk can be detected and
CC prognosis and treatment determined. See also R22559.
SC Sequence 648 AA;
SQ
Query Match 41.9%; Score 62; DB 4; Length 648;
Best Local Similarity 30.0%; Pred. No. 6.43e+01;
Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;
Db 413 tkdnlatvqcwcegslykh 432
:::|:|:|:|:|:|
1 ANDHLSLEAWSNDNDTPYLH 20
RESULT 13
ID W13107 standard; Protein; 648 AA.
AC W13107;
DT 12-MAY-1997 (first entry)
DE Human Raf-1.
KW Human; raf-1; complex; 14-3-3; beta; zeta; modulation; binding;
KW detection; screening; interaction; cell cycle; control; neoplasia;
KW pathological condition; drug.
OS Homo sapiens.
FH Key
FH Location/Qualifiers
FT 1..197
FT /note="conserved region 1 containing region"
FT 186..332

```

```

PT PN US$597719-A. /note= "conserved region 2 containing region"
PD 28-JAN-1997.
PR 14-JUL-1994; 276151.
PR 14-JUL-1994; US-276151.
PA (ONYX-) ONYX PHARM INC.
PI Freed E, Ruggieri R;
DR WPI: 97-108327/10.
DR N-PSDB; T61894.
PT Complex of raf-1 and 14-3-3 polypeptide(s) - useful for anticancer
PT drug screening
PT disclosure: Columns 31-38; 43pp; English.
CC The present sequence is human Raf-1, which can be used in a
CC novel composition comprising a complex of human Raf-1, or a
CC fragment comprising residues 1-197 or 186-332, or lacking residues
CC 51-131, and a 14-3-3 polypeptide (preferably human 14-3-3 beta or
CC zeta). The composition can be used to screen for drugs which
CC modulate the binding interaction between Raf-1 and 14-3-3.
CC especially to identify drugs that modulate Raf-1 mediated cell
CC cycle control, and/or neoplastic or other pathological conditions
CC dependent on the interaction between Raf-1 and 14-3-3 beta or zeta.
SQ Sequence 648 AA;

Query Match 41.9%; Score 62; DB 21; Length 648;
Best Local Similarity 30.0%; Pred. No. 6.43e+01;
Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

Db 413 tkdnlaivtgcggsslykh 432
::|:|:| | : | | |
Oy 1 ANDHSLILEMSDNDPYLH 20

RESULT 14
ID W17048 standard; Protein; 648 AA.
AC W17048;
DE 16-JUN-1997 (first entry)
DE Mutant mouse c-raf 1 protein used in diagnosis of cancer.
KW rat; oncogene; lymphoma; lung cancer; neoplasia; point mutation;
KW conserved region; adenocarcinoma; codon 533; diagnosis; detection.
OS Mus musculus.
PN US$518670-A.
PD 08-APR-1997.
PF 26-AUG-1988; 236947.
PF 26-AUG-1988; US-236947.
PR 16-SEP-1991; US-759738.
PR 24-JAN-1994; PR-185282.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
P1 RAPP UR, Storm SM;.
P1 WPI: 97-225421/20.
PT Classification of Lymphoma or lung cancer - on the basis of a point
PT mutation in c-raf-1 gene
PS Example 3: Column 21-24; 26pp; English.
PS W17048-W17049 are mutant versions of the mouse c-raf-1 protein, derived
PS from four different types of tumour. They were used in a method for
PS classifying a lymphoma or lung cancer, by comparison with the wild-type
PS c-raf-1 protein sequence derived from different tumours in mice. The
PS method, when used for diagnosing human tumours, involves detecting the
PS presence of a point mutation in a conserved region of the c-raf-1 gene
PS (codon 533, encoding Ser) derived from lymphoma or lung cancer tissue
PS and classifying the lymphoma or lung cancer as a c-raf-1 mutation-
PS associated cancer if one or more point mutations are present. The
PS method is particularly applicable to diagnosis of lung adenocarcinoma.
SC Sequence 648 AA;

Query Match 41.9%; Score 62; DB 22; Length 648;
Best Local Similarity 30.0%; Pred. No. 6.43e+01;
Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

Db 413 tkdnlaivtgcggsslykh 432
::|:|:| | : | | |
Oy 1 ANDHSLILEMSDNDPYLH 20
```

```

RESULT 15
ID W17049 standard; Protein; 648 AA.
AC W17049;
DT 16-JUL-1997 (first entry)
DE Mutant mouse c-raf-1 protein used in diagnosis of cancer.
KW raf; oncogene; lymphoma; lung cancer; neoplasia; point mutation;
KM conserved region; adenocarcinoma; codon 533; diagnosis; detection.
OS Mus musculus.
PN US5618670-A.
PD 08-APR-1997.
PF 26-AUG-1988; 236947.
PR 26-AUG-1988; US-236947.
PR 16-SEP-1991; US-759738.
PR 24-JAN-1994; US-185282.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI Rapp UR, Storm SM;
DR WPI: 97-225421/20.
PT Classification of lymphoma or lung cancer - on the basis of a point
  mutation in c-raf-1 gene.
  Example 3; Column 23-28; 26pp; English.
  CC W17046-W17049 are mutant versions of the mouse c-raf-1 protein, derived
  CC from four different types of tumour. They were used in a method for
  CC classifying a lymphoma or lung cancer, by comparison with the wild-type
  CC c-raf-1 protein sequence derived from different tumours in mice. The
  CC method, when used for diagnosing human tumours, involves detecting the
  CC presence of a point mutation in a conserved region of the c-raf-1 gene
  CC (codon 533, encoding Ser - this mutant version has Phe at this
  CC position) derived from lymphoma or lung cancer tissue and classifying
  CC the lymphoma or lung cancer as a c-raf-1 mutation-associated cancer if
  CC one or more point mutations are present. The method is particularly
  CC applicable to diagnosis of lung adenocarcinoma.
SQ Sequence 648 AA;

```

```

Query Match 41.9%; Score 62; DB 22; Length 648;
Best Local Similarity 30.0%; Pred. No. 6.43e+01;
Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

```

```

DB 413 tkdnlaivtgcgsslykh 432

```

```

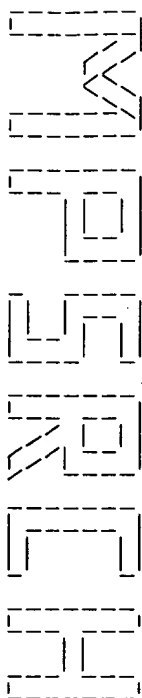
OY 1 ANDHLSILEAMSDMDTPYLH 20

```

```

Search completed: Tue Jan 11 15:36:37 2000
Job time : 37 secs.

```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MSPRCH_PP protein - protein database search, using Smith-Waterman algorithm
Run on: Tue Jan 11 15:38:57 2000; MasPar time 1.65 Seconds
151.295 Million cell updates/sec
MasPar output not generated.

Title: >US-09-290-049-1
Description: (1-21) From US09290049.pep
Perfect Score: 148
Sequence: 1 ANDHLSTLEAWSDNDPTPLMD 21

Scoring table: PAM 150
Gap 15

Searched: 122461 seqs, 11912985 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfill1

Statistics: Mean 19.740; Variance 68.890; scale 0.287

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	66	44.6	387	2	US-08-759-	Sequence 4, Applicatio	1.25e+01
2	66	44.6	464	2	US-08-759-	Sequence 22, Applicati	1.25e+01
3	63	42.6	388	2	US-08-759-	Sequence 9, Applicatio	2.41e+01
4	62	41.9	267	2	US-07-857-	Sequence 42, Applicati	3.00e+01
5	62	41.9	315	1	US-08-909-	Sequence 12, Applicati	3.00e+01
6	62	41.9	315	1	US-08-909-	Sequence 12, Applicati	3.00e+01
7	62	41.9	315	1	US-08-571-	Sequence 12, Applicati	3.00e+01
8	62	41.9	346	1	US-08-276-	Sequence 5, Applicatio	3.00e+01
9	62	41.9	648	1	US-08-185-	Sequence 12, Applicati	3.00e+01
10	62	41.9	648	2	US-08-886-	Sequence 6, Applicatio	3.00e+01
11	62	41.9	648	1	US-08-276-	Sequence 2, Applicatio	3.00e+01
12	62	41.9	648	1	US-08-185-	Sequence 3, Applicatio	3.00e+01
13	62	41.9	648	1	US-08-185-	Sequence 1, Applicatio	3.00e+01
14	62	41.9	648	1	US-08-185-	Sequence 4, Applicatio	3.00e+01
15	62	41.9	648	1	US-08-185-	Sequence 5, Applicatio	3.00e+01
16	62	41.9	648	1	US-08-185-	Sequence 2, Applicatio	3.00e+01
17	61	41.2	480	3	PCT-US95-0	Sequence 12, Applicati	3.72e+01
18	61	41.2	480	3	US-08-272-	Sequence 12, Applicati	3.72e+01
19	60	40.5	473	2	US-08-759-	Sequence 15, Applicati	4.61e+01
20	59	39.9	576	1	US-08-190-	Sequence 56, Applicati	5.71e+01
21	58	39.2	105	1	US-08-459-	Sequence 4, Applicatio	7.06e+01
22	58	39.2	770	1	US-08-525-	Sequence 1, Applicatio	7.06e+01
23	58	39.2	771	1	US-08-525-	Sequence 3, Applicatio	7.06e+01

24	57	38.5	184	2	US-08-737-	Sequence 10, Applicati	8.72e+01
25	57	38.5	500	4	5171684-2	Patent No. 5171684.	8.72e+01
26	56	37.8	161	2	US-08-403-	Sequence 22, Applicati	1.08e+02
27	56	37.8	687	1	US-08-583-	Sequence 33, Applicati	1.08e+02
28	56	37.8	687	1	US-08-164-	Sequence 31, Applicati	1.08e+02
29	56	37.8	687	1	US-08-164-	Sequence 31, Applicati	1.08e+02
30	56	37.8	687	1	US-08-583-	Sequence 31, Applicati	1.08e+02
31	56	37.8	688	1	US-08-164-	Sequence 72, Applicati	1.08e+02
32	56	37.8	688	1	US-08-164-	Sequence 72, Applicati	1.08e+02
33	56	37.8	688	1	US-08-583-	Sequence 70, Applicati	1.08e+02
34	56	37.8	688	1	US-08-583-	Sequence 72, Applicati	1.08e+02
35	56	37.8	722	3	PCT-US95-1	Sequence 2, Applicatio	1.08e+02
36	56	37.8	1651	2	US-08-447-	Sequence 2, Applicatio	1.08e+02
37	55	37.2	132	1	US-08-118-	Sequence 6, Applicatio	1.33e+02
38	55	37.2	205	2	US-08-684-	Sequence 6, Applicatio	1.33e+02
39	55	37.2	301	2	US-08-355-	Sequence 2, Applicatio	1.33e+02
40	55	37.2	301	3	PCT-US95-1	Sequence 2, Applicatio	1.33e+02
41	54	36.5	47	3	PCT-US95-0	Sequence 6, Applicatio	1.63e+02
42	54	36.5	263	2	US-08-664-	Sequence 1, Applicatio	1.63e+02
43	54	36.5	566	1	US-08-726-	Sequence 2, Applicatio	1.63e+02
44	54	36.5	695	1	US-08-583-	Sequence 8, Applicatio	1.63e+02
45	54	36.5	3898	2	US-09-059-	Sequence 2, Applicatio	1.63e+02

ALIGNMENTS

RESULT 1
ID US-08-759-581B-4 STANDARD: PRT: 387 AA.
AC xxxxxx
DT
XX
XX
XX
Sequence 4, Application US/08759581B
CC
CC Sequence 4, Application US/08759581B
CC Patent No. 5876945
CC
CC GENERAL INFORMATION:
CC APPLICANT: CHISHOLM, DEXTER A.
CC APPLICANT: DINER, BRUCE A.
CC APPLICANT: DONALDSON, GAIL K.
CC APPLICANT: HERSHEY, HOWARD P.
CC APPLICANT: JORDAN, DOUGLAS B.
CC APPLICANT: TANG, XIAO-SONG
CC APPLICANT: TROST, JEFFREY T.
CC APPLICANT: WANG, SHAOTIE
CC APPLICANT: WARREN, PATRICK V.
CC TITLE OF INVENTION: METHODS FOR IDENTIFYING HERBICIDAL AGENTS THAT INHIBIT D
CC NUMBER OF SEQUENCES: 29
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
CC STREET: 1007 MARKET STREET
CC CITY: WILMINGTON
CC STATE: DELAWARE
CC COUNTRY: U.S.A.
CC ZIP: 19898
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: DISKETTE, 3.50 INCH
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: MICROSOFT WINDOWS 3.1
CC SOFTWARE: MICROSOFT WORD 2.0C
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/759,581B
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: FLOYD, LINDA AXAMETHY
CC REGISTRATION NUMBER: 33,692
CC REFERENCE/DOCKET NUMBER: CR-9964
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 302-892-8112
CC TELEFAX: 302-773-0164
CC INFORMATION FOR SEQ ID NO: 4:

```

CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 387 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC MOLECULE TYPE: protein
CC ORIGINAL SOURCE:
CC STRAIN: SCENESMUS D1 PROTEASE PROTEIN
CC SEQUENCE 387 AA; 40578 MW; 751563 CN;

Query Match 44.6%; Score 66; DB 2; Length 387;
Best Local Similarity 36.8%; Pred. No. 1.25e+01;
Matches 7; Conservative 7; Mismatches 5; Gaps 0;

Dh 2 TSEQLFLFAMRAVDRAVY 20
Oy 1 ANDHLSILEAMSDNDTPYL 19
:::|:||||| |:|:|
LT 2
US-08-759-581B-22 STANDARD; PRT; 464 AA.
xxxxxx

Sequence 22, Application US/08759581B
DE
Sequence 22, Application US/08759581B
CC Patent No. 5876945
CC GENERAL INFORMATION:
CC APPLICANT: CHISHOLM, DEXTER A.
CC APPLICANT: DINER, BRUCE A.
CC APPLICANT: DONALDSON, GAIL K.
CC APPLICANT: HERSHEY, HOWARD P.
CC APPLICANT: JORDAN, DOUGLAS B.
CC APPLICANT: TANG, XIAO-SONG
CC APPLICANT: TROST, JEFFREY T.
CC APPLICANT: WANG, SHAOJIE
CC APPLICANT: WARREN, PATRICK V.
CC TITLE OF INVENTION: METHODS FOR IDENTIFYINGHERBICIDAL AGENTS THAT INHIBIT D1 F
CC NUMBER OF SEQUENCES: 29
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: E.I. DU PONT DE NEMOURS AND COMPANY
CC STREET: 1007 MARKET STREET
CC CITY: WILMINGTON
CC STATE: DELAWARE
CC COUNTRY: U.S.A.
CC ZIP: 19898
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: DISKETTE, 3.50 INCH
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: MICROSOFT WINDOWS 3.1
CC SOFTWARE: MICROSOFT WORD 2.0C
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/759,581B
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: FLOYD, LINDA AXAMETHY
CC REGISTRATION NUMBER: 33,692
CC REFERENCE/DOCKET NUMBER: CR-9964
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 302-892-8112
CC TELEFAX: 302-773-0164
CC INFORMATION FOR SEQ ID NO: 22:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 464 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC MOLECULE TYPE: peptide
CC SEQUENCE 464 AA; 48755 MW; 1062471 CN;

```

```

Query Match          44.6%   Score 66; DB 2; Length 464;
Best Local Similarity 36.8%; Pred. No. 1.25e+01;
Matches              7; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

Db      79 TSEQLLEAMRAVDRAVY 97
Oy      1 ANDHSLILEAMSNDMDPYL 19

RESULT    3
ID        US-08-759-SB1B-9     STANDARD; PRT; 388 AA.
AC        xxxxxx
XX
XX
DT
DE
Sequence 9, Application US/08759581B
CC
CC Sequence 9, Application US/08759581B
CC Patent No. 5876945
CC
CC GENERAL INFORMATION:
CC APPLICANT: CHISHOLM, DEXTER A.
CC APPLICANT: DINER, BROCE A.
CC APPLICANT: DONALDSON, GAIL K.
CC APPLICANT: HERSHEY, HOWARD P.
CC APPLICANT: JORDAN, DOUGLAS B.
CC APPLICANT: TANG, XIAO-SONG
CC APPLICANT: TROST, JEFFREY T.
CC APPLICANT: WANG, SHAOJIE
CC APPLICANT: WARREN, PATRICK V.
CC TITLE OF INVENTION: METHODS FOR IDENTIFYING HERBICIDAL AGENTS THAT INHIBIT D
CC NUMBER OF SEQUENCES: 29
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: E.I. DU PONT DE NEMOURS AND COMPANY
CC STREET: 1007 MARKET STREET
CC CITY: WILMINGTON
CC STATE: DELAWARE
CC COUNTRY: U.S.A.
CC ZIP: 19898
CC
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: DISKETTE, 3.50 INCH
CC
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: MICROSOFT WINDOWS 3.1
CC SOFTWARE: MICROSOFT WORD 2.0C
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/759,581B
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: FLOYD, LINDA AXAMETHY
CC REGISTRATION NUMBER: 33,692
CC REFERENCE/DOCKET NUMBER: CR-9964
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 302-773-0164
CC TELEFAX: 302-892-8112
CC INFORMATION FOR SEQ ID NO: 9:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 388 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC MOLECULE TYPE: protein
CC ORIGINAL SOURCE:
CC STRAIN: WHEAT DI PROTEINASE PROTEIN
CC
CC SEQUENCE 388 AA; 41958 MW; 750252 CN;
SQ

Query Match          42.6%   Score 63; DB 2; Length 388;
Best Local Similarity 38.9%; Pred. No. 2.41e+01;
Matches              7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Db      2 TEENLTLEAMRAVDRAVY 19
Oy      1 ANDHSLILEAMSNDMDPYL 19

```

OY 1 ANDHLSILEAMSDNDTPY 18

RESULT 4
ID US-07-857-224B-42 STANDARD: PRT; 267 AA.
XX xxxxxx

Sequence 42, Application US/07857224B
Sequence 42, Application US/07857224B
Patent No. 5958784
GENERAL INFORMATION:
APPLICANT: Benner, Steven A.
TITLE OF INVENTION: Predicting Folded Structures of Proteins
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:
ADDRESSEE: Steven A. Benner
STREET: Hadlaubstrasse 151
CITY: Zurich
STATE: none
COUNTRY: Switzerland
ZIP: (note: this is an international post code) CH-8092
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch diskette, 1.4 MB storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/857, 224B
FILING DATE: 03/25/92
CLASSIFICATION: 436
PRIOR APPLICATION DATA: none
TELECOMMUNICATION INFORMATION:
TELEPHONE: (international) 41 1 632 2830
TELEFAX: (international) 41 1 262 2437
TELEX: none
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 267
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: protein
ORIGINAL SOURCE:
ORGANISM: Saccharomyces cerevisiae
FEATURE: Protein kinase; Table 8 Column 48
PUBLICATION INFORMATION:
AUTHORS:
AUTHORS: Hanks, S. K.
AUTHORS: Quinn, A. M.
AUTHORS: Hunter, T.
TITLE: The protein kinase family
JOURNAL: Science
VOLUME: 241
PAGES: 42-52
DATE: 1988
SEQUENCE 267 AA; 30518 MW; 382200 CN;

Query Match 41.9%; Score 62; DB 2; Length 267;
Best Local Similarity 30.0%; Pred. No. 3.00e+01;
Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

DB 68 TKDNLAIYTWCEGSSLYKH 87
OY 1 ANDHLSILEAMSDNDTPYLH 20

RESULT 5
ID US-08-909-984A-12 STANDARD: PRT; 315 AA.
XX

AC xxxxxx
XX
DT
DE
XX
XX
Sequence 12, Application US/08909984A
Patent No. 5747275
GENERAL INFORMATION:
APPLICANT: Rubin, Gerry M.
APPLICANT: Therrien, Marc
APPLICANT: Chang, Henry C.
APPLICANT: Karim, Felix D.
APPLICANT: Wasserman, David A.
TITLE OF INVENTION: A No. 5747275e1 Protein Kinase Required for Ras
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/909, 984A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A.
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: B96-010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 315 amino acids
TYPE: amino acid
TOPOLOGY: not relevant
STRANDEDNESS: not relevant
MOLECULE TYPE: peptide
SEQUENCE 315 AA; 35949 MW; 526392 CN;

Query Match 41.9%; Score 62; DB 1; Length 315;
Best Local Similarity 30.0%; Pred. No. 3.00e+01;
Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

DB 80 TKDNLAIYTWCEGSSLYKH 99
OY 1 ANDHLSILEAMSDNDTPYLH 20

RESULT 6
ID US-08-909-983-12 STANDARD: PRT; 315 AA.
XX
XX
AC xxxxxx
XX
DT
DE
XX
XX
Sequence 12, Application US/08909983
Patent No. 5747288
GENERAL INFORMATION:
APPLICANT: Rubin, Gerry M.
APPLICANT: Therrien, Marc
APPLICANT: Chang, Henry C.

```

CC APPLICANT: Karim, Felix D.
CC APPLICANT: Massarman, David A.
CC TITLE OF INVENTION: A No. 5747288e1 Protein Kinase Required for Ras
CC TITLE OF INVENTION: Signal Transduction
CC NUMBER OF SEQUENCES: 12
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
CC STREET: 268 BUSH STREET, SUITE 3200
CC CITY: SAN FRANCISCO
CC STATE: CALIFORNIA
CC COUNTRY: USA
CC ZIP: 94104
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/909,983
CC FILING DATE: 12-JUN-1997
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/571,758
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: OSMAN, RICHARD A.
CC REGISTRATION NUMBER: 36,627
CC REFERENCE/DOCKET NUMBER: B96-010
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 343-4341
CC TELEFAX: (415) 343-4342
CC INFORMATION FOR SRD ID NO: 12:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 315 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: not relevant
CC TOPOLOGY: not relevant
CC MOLECULE TYPE: peptide
CC SEQUENCE 315 AA; 35949 MM; 526392 CN;
SO

Query Match          41.9%; Score 62; DB 1; Length 315;
Best Local Similarity 30.0%; Pred. NO. 3.00e+01;
Matches      6; Conservative    7; Mismatches    7; Indels    0; Gaps    0;

Db      80 TKDNLAIVTOWCEGSSLYKH 99
       ::|||:: ||::|||
O      1 ANDHLSILEAWSDDNDTPYLH 20

RESULT   7
ID US-08-571-758-12 STANDARD: PRT: 315 AA.
XX
XX xxxxxx
XX
XX Sequence 12, Application US/08571758
DE
CC Sequence 12, Application US/08571758
CC Patent No. 5700675
CC GENERAL INFORMATION:
CC APPLICANT: Rubin, Gerry M.
CC APPLICANT: Therrien, Marc
CC APPLICANT: Chang, Henry C.
CC APPLICANT: Karim, Felix D.
CC APPLICANT: Massarman, David A.
CC TITLE OF INVENTION: A No. 5700675e1 Protein Kinase Required for Ras
CC TITLE OF INVENTION: Signal Transduction
CC NUMBER OF SEQUENCES: 12
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
CC STREET: 268 BUSH STREET, SUITE 3200
CC CITY: SAN FRANCISCO

```

CC	STATE:	CALIFORNIA
CC	COUNTRY:	USA
CC	ZIP:	94104
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE:	Floppy disk
CC	COMPUTER:	IBM PC compatible
CC	OPERATING SYSTEM:	PC-DOS/MS-DOS
CC	SOFTWARE:	Patentin Release #1.0, Version #1.30
CC	CURRENT APPLICATION DATA:	
CC	APPLICATION NUMBER:	US/08/571,758
CC	FILING DATE:	
CC	CLASSIFICATION:	435
CC	ATTORNEY/AGENT INFORMATION:	
CC	NAME:	OSMAN, RICHARD A.
CC	REGISTRATION NUMBER:	36,627
CC	REFERENCE/DOCKET NUMBER:	B96-010
CC	TELECOMMUNICATION INFORMATION:	
CC	TELEPHONE:	(415) 343-4341
CC	TELEFAX:	(415) 343-4342
CC	INFORMATION FOR SEO ID NO:	12:
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH:	315 amino acids
CC	TYPE:	amino acid
CC	STRANDEDNESS:	not relevant
CC	TOPOLOGY:	not relevant
CC	MOLECULE TYPE:	peptide
SQ	SEQUENCE	315 AA; 35949 MW; 526392 CN;
Db	Query Match	41.9%; Score 62; DB 1; Length 315;
Oy	Best Local Similarity	30.0%; Pred. No. 3.00e+01;
	Matches	6; Conservative 7; Mismatches 7; Indels 0; Gaps 0
Db	80 TKDMLAIVTWCESGSLYKH 99	
Oy	1 ANDHLSTLEAWSNDNTPRYH 20	
RESULT	8	
ID	US-08-276-151-5	STANDARD; PRT; 346 AA.
AC	xxxxxx	
XX	Sequence 5, Application US/08276151	
DE	Sequence 5, Application US/08276151	
CC	Patent No. 5597719	
CC	GENERAL INFORMATION:	
CC	APPLICANT:	Freed, Ellen
CC	APPLICANT:	Ruggieri, Rosamaría
CC	TITLE OF INVENTION:	Interaction of raf-1 and 14-3-3 Proteins
CC	NUMBER OF SEQUENCES:	9
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESSEE:	Cooley Godward et al.
CC	STREET:	Five Palo Alto Square
CC	CITY:	Palo Alto
CC	STATE:	CA
CC	COUNTRY:	USA
CC	ZIP:	94036
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE:	Floppy disk
CC	COMPUTER:	IBM PC compatible
CC	OPERATING SYSTEM:	PC-DOS/MS-DOS
CC	SOFTWARE:	Patentin Release #1.0, Version #1.25
CC	CURRENT APPLICATION DATA:	
CC	APPLICATION NUMBER:	US/08/276,151
CC	FILING DATE:	14-JUL-1994
CC	CLASSIFICATION:	530
CC	ATTORNEY/AGENT INFORMATION:	
CC	NAME:	Torchia, Ph.D., Timothy E
CC	REGISTRATION NUMBER:	36,700
CC	REFERENCE/DOCKET NUMBER	ONYX-005/0005

SEQUENCE	AA	73051 MW	2254837 CN
Query Match	41.9%	Score 62;	DB 1; Length 648;
Best Local Similarity	30.0%	Pred. No. 3.00e+01;	
Matches	6;	Conservative	7; Mismatches 7; Indels 0; Gaps 0;
Db	413 TKDNLAIYTCWCGSSLYKH	432	
QY	1 ANDHSLILEAWSMDNDPYLH	20	
RESULT	10		
ID	US-08-886-751A-6	STANDARD;	PRT; 648 AA.
AC	xxxxxx		
DT			
XX			
Sequence	6, Application US/0886751A		
Sequence	6, Application US/0886751A		
Patent No.	5865783		
GENERAL INFORMATION:			
APPLICANT:	YOO, Tai-June		
APPLICANT:	Cheng, Kuang-Chuan		
TITLE OF INVENTION:	Autoimmune Inner Ear Disease Antigen and		
TITLE OF INVENTION:	Diagnostic Assay		
NUMBER OF SEQUENCES:	6		
CORRESPONDENCE ADDRESS:			
ADDRESSEE:	Jones & Askew		
STREET:	191 Peachtree Street, 37th Floor		
CITY:	Atlanta		
STATE:	Georgia		
COUNTRY:	U.S.A.		
ZIP:	30303		
COMPUTER READABLE FORM:			
MEDIUM TYPE:	Floppy disk		
COMPUTER:	IBM PC compatible		
OPERATING SYSTEM:	PC-DOS/MS-DOS		
SOFTWARE:	Patent Release #1.0, Version #1.30		
CURRENT APPLICATION DATA:			
APPLICATION NUMBER:	US/08/886,751A		
FILING DATE:			
CLASSIFICATION:	424		
ATTORNEY/AGENT INFORMATION:			
NAME:	Young, Leona G.		
REGISTRATION NUMBER:	37,266		
REFERENCE/DOCKET NUMBER:	25490-0100		
TELECOMMUNICATION INFORMATION:			
TELEPHONE:	(404) 818-3700		
TELEFAX:	(404) 818-3799		
INFORMATION FOR SEQ ID NO:	6:		
SEQUENCE CHARACTERISTICS:			
LENGTH:	648 amino acids		
TYPE:	amino acid		
STRANDEDNESS:	single		
TOPOLOGY:	linear		
MOLECULE TYPE:	protein		
HYPOTHETICAL:	NO		
ANTI-SENSE:	NO		
FRAGMENT TYPE:	N-terminal		
ORIGINAL SOURCE:			
ORGANISM:	Homo sapiens		
SEQUENCE	648 AA; 73051 MW; 2254837 CN;		
Query Match	41.9%	Score 62;	DB 2; Length 648;
Best Local Similarity	30.0%	Pred. No. 3.00e+01;	
Matches	6;	Conservative	7; Mismatches 7; Indels 0; Gaps 0;
Db	413 TKDNLAIYTCWCGSSLYKH	432	
QY	1 ANDHSLILEAWSMDNDPYLH	20	

CC INFORMATION FOR SEQ ID NO: 4:

Db 413 TKDNLAIVTQWCEGSSLYKH 432

QY 1 ANDHLSILEAWSNDNDTPYLH 20

Thu Jan 13 09:13:31 2000

US-09-290-049-1.rai

Search completed: Tue Jan 11 15:39:05 2000
Job time : 8 secs.
